# Supplement to
# INFERENCE BASED ON CONDITIONAL MOMENT INEQUALITIES

**By**

**Donald W.K. Andrews and Xiaoxia Shi**

**June 2010**
**Revised July 2011**

**COWLES FOUNDATION DISCUSSION PAPER NO. 1761R**

**COWLES FOUNDATION FOR RESEARCH IN ECONOMICS**
**YALE UNIVERSITY**
**Box 208281**
**New Haven, Connecticut 06520-8281**

**http://cowles.econ.yale.edu/**

# Supplemental Material

for

# Inference Based on Conditional Moment Inequalities

Donald W. K. Andrews
Cowles Foundation for Research in Economics
Yale University

Xiaoxia Shi
Department of Economics
University of Wisconsin, Madison

November 2007
Revised: July 2011

# Contents

# 11 Outline

This Supplement includes six appendices.

Supplemental Appendix A gives proofs of Theorems 1 and 2(a).

Supplemental Appendix B provides a number of supplemental results to the main paper. These include:

(i) results for Kolmogorov-Smirnov (KS) and approximate Cramér von Mises (A-CvM) tests and CS's in Section 13.1,

(ii) three additional examples of collections $\mathcal{G}$ and probability measures $Q$ that satisfy Assumptions CI, M, FA(e), and Q in Section 13.2,

(iii) an illustration of the verification of Assumptions LA1-LA3 in Section 13.3,

(iv) an illustration of some uniformity issues that arise with infinite-dimensional nuisance parameters in Section 13.4,

(v) an illustration of problems with pointwise asymptotics in Section 13.5, and

(vi) coverage probability results for subsampling tests and CS's under drifting sequences of distributions in Section 13.6.

Supplemental Appendix C provides proofs of the results that are stated in the main paper but are not proved in Supplemental Appendix A. These include:

(i) the proofs of Lemmas 2 and 3 and Theorem 2(b) in Section 14.1,

(ii) the proofs of Lemma 4 and Theorem 3 concerning fixed alternatives in Section 14.2,

(iii) the proof of Theorem 4 concerning local power in Section 14.3, and

(iv) the proof of Lemma 1 concerning the verification of Assumptions S1-S4 in Section 14.4.

Supplemental Appendix D provides proofs of the results stated in Supplemental Appendix B. These include:

(i) the proofs of Kolmogorov-Smirnov and approximate Cramér von Mises results in Section 15.1,

(ii) the proof of Lemma B2 in Section 15.2,

(iii) the proofs of Theorems B4 and B5 regarding uniformity issues in Section 15.3, and

(iv) the proofs of the subsampling results in Section 15.4.

Supplemental Appendix E proves Lemma A1, which is stated in Supplemental Appendix A.

Supplemental Appendix F provides the simulation results for the mean selection and interval-outcome regression models and additional material (and results) concerning the simulations in the quantile selection and entry game models.

# 12 Supplemental Appendix A

In this Appendix, we provides proofs of the uniform asymptotic coverage probability results for GMS and PA CS's. In particular, it proves Theorems 1 and 2(a). Proofs of the other results stated in the paper are given in Supplemental Appendix C.

## 12.1 Proof of Theorem 1

The following Lemma is used in the proofs of Theorems 1, 2, 3, and 4. It establishes a functional CLT and uniform LLN for certain independent non-identically distributed empirical processes.

Let $h_2$ denote a $k \times k$-matrix-valued covariance kernel on $\mathcal{G} \times \mathcal{G}$ (such as an element of $\mathcal{H}_2$).

**Definition SubSeq($\mathbf{h_2}$).** $SubSeq(h_2)$ is the set of subsequences $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\}$, where $\{a_n : n \geq 1\}$ is some subsequence of $\{n\}$, for which

$$\text{(i)} \lim_{n \to \infty} \sup_{g, g^* \in \mathcal{G}} ||h_{2,F_{a_n}}(\theta_{a_n}, g, g^*) - h_2(g, g^*)|| = 0,$$

(ii) $\theta_{a_n} \in \Theta$, (iii) $\{W_i : i \geq 1\}$ are i.i.d. under $F_{a_n}$, (iv) $Var_{F_{a_n}}(m_j(W_i, \theta_{a_n})) > 0$ for $j = 1, ..., k$, for $n \geq 1$, (v) $\sup_{n \geq 1} E_{F_{a_n}} |m_j(W_i, \theta_{a_n}) / \sigma_{F_{a_n}, j}(\theta_{a_n})|^{2+\delta} < \infty$ for $j = 1, ..., k$, for some $\delta > 0$, and (vi) Assumption M holds with $F_{a_n}$ in place of $F$ and $F_n$ in Assumptions M(b) and M(c), respectively.

The sample paths of the Gaussian process $\nu_{h_2}(\cdot)$, which is defined in (4.2) and appears in the following Lemma, are bounded and uniformly $\rho$-continuous a.s. The pseudo-metric $\rho$ on $\mathcal{G}$ is a pseudo-metric commonly used in the empirical process literature:

$$\rho^2(g, g^*) = tr\left(h_2(g, g) - h_2(g, g^*) - h_2(g^*, g) + h_2(g^*, g^*)\right). \tag{12.1}$$

For $h_2(\cdot, \cdot) = h_{2,F}(\theta, \cdot, \cdot)$, where $(\theta, F) \in \mathcal{F}$, this metric can be written equivalently as

$$\begin{aligned} \rho^2(g, g^*) &= E_F ||D_F^{-1/2}(\theta)[\widetilde{m}(W_i, \theta, g) - \widetilde{m}(W_i, \theta, g^*)]||^2, \text{ where} \\ \widetilde{m}(W_i, \theta, g) &= m(W_i, \theta, g) - E_F m(W_i, \theta, g). \end{aligned} \tag{12.2}$$

**Lemma A1.** *For any subsequence* $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\} \in SubSeq(h_2)$,
(a) $\nu_{a_n, F_{a_n}}(\theta_{a_n}, \cdot) \Rightarrow \nu_{h_2}(\cdot)$ *as* $n \to \infty$ (*as processes indexed by* $g \in \mathcal{G}$), *and*

(b) $\sup_{g,g^* \in \mathcal{G}} ||\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n}, g, g^*) - h_2(g, g^*)|| \rightarrow_p 0$ *as* $n \rightarrow \infty$.

**Comments. 1.** The proof of Lemma A1 is given in Supplemental Appendix E. Part (a) is proved by establishing the manageability of $\{m(W_i, \theta_{a_n}, g) - E_{F_{a_n}} m(W_i, \theta_{a_n}, g) : g \in \mathcal{G}\}$ and by establishing a functional CLT for $R^k$-valued i.n.i.d. empirical processes with the pseudo-metric $\rho$ by using the functional CLT in Pollard (1990, Thm. 10.2) for real-valued empirical processes. Part (b) is proved using a maximal inequality given in Pollard (1990, (7.10)).

**2.** To obtain uniform asymptotic coverage probability results for CS's, Lemma A1 is applied with $(\theta_{a_n}, F_{a_n}) \in \mathcal{F}$ for all $n \geq 1$ and $h_2 \in \mathcal{H}_2$. In this case, conditions (ii)-(vi) in the definition of $SubSeq(h_2)$ hold automatically by the definition of $\mathcal{F}$. To obtain power results under fixed and local alternatives, Lemma A1 is applied with $(\theta_{a_n}, F_{a_n}) \notin \mathcal{F}$ for all $n \geq 1$ and $h_2$ may or may not be in $\mathcal{H}_2$.

**Proof of Theorem 1.** First, we prove part (a). Let $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ be a sequence for which $h_{2,F_n}(\theta_n) \in \mathcal{H}_{2,cpt}$ for all $n \geq 1$ and the term in square brackets in Theorem 1(a) evaluated at $(\theta_n, F_n)$ differs from its supremum over $(\theta, F) \in \mathcal{F}$ with $h_{2,F}(\theta) \in \mathcal{H}_{2,cpt}$ by $\delta_n$ or less, where $0 < \delta_n \rightarrow 0$ as $n \rightarrow \infty$. Such a sequence always exists. To prove part (a), it suffices to show that part (a) holds with the supremum deleted and with $(\theta, F)$ replaced by $(\theta_n, F_n)$.

By the compactness of $\mathcal{H}_{2,cpt}$, given any subsequence $\{u_n : n \geq 1\}$ of $\{n\}$, there exists a subsubsequence $\{a_n : n \geq 1\}$ for which $d(h_{2,F_{a_n}}(\theta_{a_n}), h_{2,0}) \rightarrow 0$ as $n \rightarrow \infty$ for some $\theta_0 \in \Theta$, where $d$ is defined in (5.6), and some $h_{2,0} \in \mathcal{H}_{2,cpt}$. This and $(\theta_{a_n}, F_{a_n}) \in \mathcal{F}$ for all $n \geq 1$ implies that $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\} \in SubSeq(h_{2,0})$.

Now, by Lemma A1, we have

$$\begin{pmatrix} \nu_{a_n,F_{a_n}}(\theta_{a_n}, \cdot) \\ \widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n}, \cdot) \end{pmatrix} \Rightarrow \begin{pmatrix} \nu_{h_{2,0}}(\cdot) \\ h_{2,0}(\cdot) \end{pmatrix} \text{ as } n \rightarrow \infty \tag{12.3}$$

as stochastic processes on $\mathcal{G}$, where $\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n}, g) = \widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n}, g, g)$ and $h_{2,0}(g) = h_{2,0}(g, g)$.

Given this, by the almost sure representation theorem, e.g., see Pollard (1990, Thm. 9.4), there exists a probability space and random quantities $\tilde{\nu}_{a_n}(\cdot)$, $\tilde{h}_{2,a_n}(\cdot)$, $\tilde{\nu}_0(\cdot)$, and $\tilde{h}_2(\cdot)$ defined on it such that (i) $(\tilde{\nu}_{a_n}(\cdot), \tilde{h}_{2,a_n}(\cdot))$ has the same distribution as $(\nu_{a_n,F_{a_n}}(\theta_{a_n}, \cdot), \widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n}, \cdot))$, (ii) $(\tilde{\nu}_0(\cdot), \tilde{h}_2(\cdot))$ has the same distribution as $(\nu_{h_{2,0}}(\cdot),$

$h_{2,0}(\cdot))$, and

$$\text{(iii)} \sup_{g\in\mathcal{G}} \left\| \begin{pmatrix} \tilde{\nu}_{a_n}(g) \\ \tilde{h}_{2,a_n}(g) \end{pmatrix} - \begin{pmatrix} \tilde{\nu}_0(g) \\ \tilde{h}_2(g) \end{pmatrix} \right\| \to 0 \text{ as } n\to\infty \text{ a.s.} \tag{12.4}$$

Because $h_{2,0}(\cdot)$ is deterministic, condition (ii) implies that $\tilde{h}_2(\cdot) = h_{2,0}(\cdot)$ a.s.

Define

$$\begin{aligned}
\tilde{h}^{\varepsilon}_{2,a_n}(\cdot) &= \tilde{h}_{2,a_n}(\cdot) + \varepsilon \cdot Diag(\tilde{h}_{2,a_n}(1_k)), \\
\tilde{T}_{a_n} &= \int S(\tilde{\nu}_{a_n}(g) + h_{1,a_n,F_{a_n}}(\theta_{a_n}, g), \tilde{h}^{\varepsilon}_{2,a_n}(g))dQ(g), \\
h^{\varepsilon}_{2,0}(\cdot) &= h_{2,0}(\cdot) + \varepsilon I_k, \text{ and} \\
\tilde{T}_{a_n,0} &= \int S\left(\tilde{\nu}_0(g) + h_{1,a_n,F_{a_n}}(\theta_{a_n}, g), h^{\varepsilon}_{2,0}(g)\right) dQ(g).
\end{aligned} \tag{12.5}$$

By construction, $\tilde{T}_{a_n}$ and $T_{a_n}(\theta_{a_n})$ have the same distribution, and $\tilde{T}_{a_n,0}$ and $T(h_{a_n,F_{a_n}}(\theta_{a_n}))$ have the same distribution for all $n \geq 1$.

Hence, to prove part (a), it suffices to show that

$$A = \limsup_{n\to\infty} \left[P_{F_{a_n}}(\tilde{T}_{a_n} > x_{h_{a_n,F_{a_n}}(\theta_{a_n})}) - P(\tilde{T}_{a_n,0} + \delta > x_{h_{a_n,F_{a_n}}(\theta_{a_n})})\right] \leq 0. \tag{12.6}$$

Below we show that

$$\tilde{T}_{a_n} - \tilde{T}_{a_n,0} \to 0 \text{ as } n \to \infty \text{ a.s.} \tag{12.7}$$

Let

$$\begin{aligned}
\widetilde{\Delta}_n &= 1(\tilde{T}_{a_n,0} + (\tilde{T}_{a_n} - \tilde{T}_{a_n,0}) > x_{h_{a_n,F_{a_n}}(\theta_{a_n})}) - 1(\tilde{T}_{a_n,0} + \delta > x_{h_{a_n,F_{a_n}}(\theta_{a_n})}) \\
&= \widetilde{\Delta}^+_n - \widetilde{\Delta}^-_n, \text{ where} \\
\widetilde{\Delta}^+_n &= \max\{\widetilde{\Delta}_n, 0\} \in [0,1] \text{ and } \widetilde{\Delta}^-_n = \max\{-\widetilde{\Delta}_n, 0\} \in [0,1].
\end{aligned} \tag{12.8}$$

By (12.7) and $\delta > 0$, $\lim_{n\to\infty} \widetilde{\Delta}^+_n = 0$ a.s. Hence, by the bounded convergence theorem,

$$\begin{aligned}
\lim_{n\to\infty} E_{F_{a_n}} \widetilde{\Delta}^+_n &= 0 \text{ and} \\
A = \limsup_{n\to\infty} E_{F_{a_n}} \widetilde{\Delta}_n &= \limsup_{n\to\infty} E_{F_{a_n}} \widetilde{\Delta}^+_n - \liminf_{n\to\infty} E_{F_{a_n}} \widetilde{\Delta}^-_n \\
&= -\liminf_{n\to\infty} E_{F_{a_n}} \widetilde{\Delta}^-_n \leq 0.
\end{aligned} \tag{12.9}$$

Hence, (12.6) holds and the proof of part (a) is complete, except for (12.7).

To prove part (b), analogous results to (12.6), (12.8), and (12.9) hold by analogous arguments.

It remains to show (12.7). We do so by fixing a sample path $\omega$ and using the bounded convergence theorem (because $\tilde{T}_{a_n}$ and $\tilde{T}_{a_n,0}$ are both integrals over $g \in \mathcal{G}$ with respect to the measure $Q$). Let $\tilde{\Omega}$ be the collection of all $\omega \in \Omega$ such that $(\tilde{\nu}_{a_n}(g), \tilde{h}_{2,a_n}(g))(\omega)$ converges to $(\tilde{\nu}_0(g), h_{2,0}(g))(\omega)$ uniformly over $g \in \mathcal{G}$ as $n \to \infty$ and $\sup_{g \in \mathcal{G}} \|\tilde{\nu}_0(g)(\omega)\| < \infty$. By (12.4) and $\tilde{h}_2(\cdot) = h_{2,0}(\cdot)$ a.s., $P(\tilde{\Omega}) = 1$. Consider a fixed $\omega \in \tilde{\Omega}$. By Assumption S2 and (12.4), for all $g \in \mathcal{G}$,

$$\sup_{\mu \in [0,\infty)^p \times \{0\}^v} \left| S\left(\tilde{\nu}_{a_n}(g)(\omega) + \mu, \tilde{h}^{\varepsilon}_{2,a_n}(g)(\omega)\right) - S\left(\tilde{\nu}_0(g)(\omega) + \mu, h^{\varepsilon}_{2,0}(g)\right)\right| \to 0 \quad (12.10)$$

as $n \to \infty$ a.s. Thus, for all $g \in \mathcal{G}$ and all $\omega \in \widetilde{\Omega}$,

$$\begin{aligned} & S\left(\tilde{\nu}_{a_n}(g)(\omega) + h_{1,a_n,F_{a_n}}(\theta_{a_n}, g), \tilde{h}^{\varepsilon}_{2,a_n}(g)(\omega)\right) \\ & \quad - S\left(\tilde{\nu}_0(g)(\omega) + h_{1,a_n,F_{a_n}}(\theta_{a_n}, g), h^{\varepsilon}_{2,0}(g)\right) \\ & \to 0 \text{ as } n \to \infty. \end{aligned} \quad (12.11)$$

Next, we show that for fixed $\omega \in \tilde{\Omega}$ the first summand on the left-hand side of (12.11) is bounded by a constant. Let $0 < \chi < 1$. By (12.4), there exists $N < \infty$ such that for all $n \geq N$,

$$\sup_{g \in \mathcal{G}} \|\tilde{\nu}_{a_n}(g)(\omega) - \tilde{\nu}_0(g)(\omega)\| < \chi \text{ and } \left\| Diag(\tilde{h}_{2,a_n}(1_k))(\omega) - I_k \right\| < \chi \quad (12.12)$$

using the fact that $Diag(h_{2,0}(1_k)) = I_k$ by construction. Let $B_\chi(\omega) = \sup_{g \in \mathcal{G}} ||\tilde{\nu}_0(g)(\omega)|| + \chi$. Then, for all $n \geq N$,

$$\sup_{g \in \mathcal{G}} \|\tilde{\nu}_{a_n}(g)(\omega)\| \leq B_\chi(\omega) < \infty. \quad (12.13)$$

First, consider the case where no moment equalities are present, i.e., $v = 0$ and

$k = p$. In this case, for $n \geq N$, we have: for all $g \in \mathcal{G}$,

$$
\begin{aligned}
0 &\leq S(\tilde{\nu}_{a_n}(g)(\omega) + h_{1,a_n,F_{a_n}}(\theta_{a_n}, g), \tilde{h}^{\varepsilon}_{2,a_n}(g)(\omega)) \\
&\leq S(\tilde{\nu}_{a_n}(g)(\omega), \tilde{h}^{\varepsilon}_{2,a_n}(g)(\omega)) \\
&\leq S(-B_{\chi}(\omega)1_p, \varepsilon \cdot Diag(\tilde{h}_{2,a_n}(1_p))) \\
&\leq S(-B_{\chi}(\omega)1_p, \varepsilon(1-\chi)I_p), \qquad (12.14)
\end{aligned}
$$

where the first inequality holds by Assumption S1(c), the second inequality holds by Assumption S1(b) and $h_{1,a_n,F_{a_n}}(\theta_{a_n}, g) \geq 0_p$ (which holds because $(\theta_{a_n}, F_{a_n}) \in \mathcal{F}$), the third inequality holds by Assumption S1(b) and (12.13) as well as by Assumption S1(e) and the definition of $\tilde{h}^{\varepsilon}_{2,a_n}(g)(\omega)$ in (12.5), and the last inequality holds by Assumption S1(e) and (12.12). For fixed $\omega \in \tilde{\Omega}$, the constant $S(-B_{\chi}(\omega)1_p, \varepsilon(1-\chi)I_p)$ bounds the first summand on the left-hand side of (12.11) for all $n \geq N$.

For the case where $v > 0$, the third inequality in (12.14) needs to be altered because $S(m, \Sigma)$ is not assumed to be non-increasing in $m_{II}$, where $m = (m_I', m_{II}')'$. In this case, for the bound with respect to the last $v$ elements of $\tilde{\nu}_{a_n}(g)(\omega)$, denoted by $\tilde{\nu}_{a_n,II}(g)(\omega)$, we use the continuity condition on $S(m, \Sigma)$, i.e., Assumption S1(d), which yields uniform continuity of $S(-B_{\chi}(\omega)1_p, m_{II}, \varepsilon(1-\chi)I_k)$ over the compact set $\{m_{II} : ||m_{II}|| \leq B_{\chi}(\omega) < \infty\}$ and delivers a finite bound because $\sup_{g \in \mathcal{G}, n \geq 1} ||\tilde{\nu}_{a_n,II}(g)(\omega)|| \leq B_{\chi}(\omega)$.

By an analogous but simpler argument, for fixed $\omega \in \tilde{\Omega}$, the second summand on the left-hand side of (12.11) is bounded by a constant.

Hence, the conditions of the bounded convergence theorem hold and for fixed $\omega \in \tilde{\Omega}$, $\tilde{T}_{a_n}(\omega) - \tilde{T}_{a_n,0}(\omega) \to 0$ as $n \to \infty$. Thus, (12.7) holds and the proof is complete. $\square$

## 12.2 Proof of Theorem 2(a)

For GMS CS's, Theorem 2(a) follows immediately from the following three Lemmas. The PA critical value is a GMS critical value with $\varphi_n(x) = 0$ for all $x \in R$ and this function $\varphi_n(x)$ satisfies Assumption GMS1 (though not Assumption GMS2(b)). Hence, Theorem 2(a) for GMS CS's covers PA CS's.

**Lemma A2**. *Suppose Assumptions* M, S1, *and* S2 *hold. Then, for every compact* *subset* $\mathcal{H}_{2,cpt}$ *of* $\mathcal{H}_2$ *and all* $\delta > 0$,

$$\limsup_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}:\\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} P_F\left(T_n(\theta) > c_0(h_{n,F}(\theta), 1-\alpha) + \delta\right) \le \alpha.$$

**Lemma A3**. *Suppose Assumptions* M, S1, *and* GMS1 *hold. Then, for every compact subset* $\mathcal{H}_{2,cpt}$ *of* $\mathcal{H}_2$,

$$\lim_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}:\\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} P_F\left(c(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha) < c(h_{1,n,F}(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha)\right) = 0.$$

**Lemma A4.** *Suppose Assumptions* M, S1, *and* S2 *hold. Then, for every compact subset* $\mathcal{H}_{2,cpt}$ *of* $\mathcal{H}_2$ *and for all* $0 < \delta < \eta$ *(where* $\eta$ *is as in the definition of* $c(h, 1-\alpha)$*),*

$$\lim_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}:\\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} P_F\Big(c(h_{1,n,F}(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha) < c_0(h_{1,n,F}(\theta), h_{2,F}(\theta), 1-\alpha) + \delta\Big) = 0.$$

The following Lemma is used in the proof of Lemma A4.

**Lemma A5.** *Suppose Assumptions* M, S1, *and* S2 *hold. Let* $\{h_{2,n} : n \ge 1\}$ *and* $\{h^*_{2,n} : n \ge 1\}$ *be any two sequences of* $k \times k$*-valued covariance kernels on* $\mathcal{G} \times \mathcal{G}$ *such that* $d(h_{2,n}, h^*_{2,n}) \to 0$ *and* $d(h_{2,n}, h_{2,0}) \to 0$ *for some* $k \times k$*-valued covariance kernel* $h_{2,0}$ *on* $\mathcal{G} \times \mathcal{G}$*. Then, for all* $\eta_1 > 0$ *and all* $\delta > 0$,

$$\liminf_{n\to\infty} \inf_{h_1\in\mathcal{H}_1} \left[c_0(h_1, h_{2,n}, 1-\alpha+\eta_1) + \delta - c_0(h_1, h^*_{2,n}, 1-\alpha)\right] \ge 0.$$

**Proof of Lemma A2.** For all $\delta > 0$, we have

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}:\\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} P_F\left(T_n(\theta) > c_0(h_{n,F}(\theta), 1-\alpha) + \delta\right) \\
&\le \limsup_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}:\\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} \left[P_F\left(T_n(\theta) > c_0(h_{n,F}(\theta), 1-\alpha) + \delta\right)\right. \\
&\qquad\qquad \left. - P\left(T(h_{n,F}(\theta)) > c_0(h_{n,F}(\theta), 1-\alpha)\right)\right] \\
&\quad + \limsup_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}:\\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} P\left(T(h_{n,F}(\theta)) > c_0(h_{n,F}(\theta), 1-\alpha)\right) \\
&\le 0 + \alpha,
\end{aligned} \tag{12.15}$$

where the second inequality holds by Theorem 1(a) with $x_{h_{n,F}(\theta)} = c_0(h_{n,F}(\theta), 1-\alpha)+\delta$ and by the definition of the quantile $c_0(h_{n,F}(\theta), 1-\alpha)$ of $T(h_{n,F}(\theta))$. $\square$

**Proof of Lemma A3.** Let $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ be a sequence for which $h_{2,F_n}(\theta_n) \in \mathcal{H}_{2,cpt}$ and the probability in the statement of the Lemma evaluated at $(\theta_n, F_n)$ differs from its supremum over $(\theta, F) \in \mathcal{F}$ (with $h_{2,F}(\theta) \in \mathcal{H}_{2,cpt}$) by $\delta_n$ or less, where $0 < \delta_n \to 0$ as $n \to \infty$. Such a sequence always exists. It suffices to show

$$\lim_{n\to\infty} P_{F_n}\left(c(\varphi_n(\theta_n), \widehat{h}_{2,n}(\theta_n), 1-\alpha) < c(h_{1,n,F_n}(\theta_n), \widehat{h}_{2,n}(\theta_n), 1-\alpha)\right) = 0. \tag{12.16}$$

By the compactness of $\mathcal{H}_{2,cpt}$, given any subsequence $\{u_n : n \geq 1\}$ of $\{n\}$, there exists a subsubsequence $\{a_n : n \geq 1\}$ for which $d(h_{2,F_{a_n}}(\theta_{a_n}), h_{2,0}) \to 0$ as $n \to \infty$ for some $h_{2,0} \in \mathcal{H}_{2,cpt}$. This and $(\theta_{a_n}, F_{a_n}) \in \mathcal{F}$ for all $n \geq 1$ implies that $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\} \in SubSeq(h_{2,0})$. Hence, it suffices to show

$$\lim_{n\to\infty} P_{F_{a_n}}\left(c(\varphi_{a_n}(\theta_{a_n}), \widehat{h}_{2,a_n}(\theta_{a_n}), 1-\alpha) < c(h_{1,a_n,F_{a_n}}(\theta_{a_n}), \widehat{h}_{2,a_n}(\theta_{a_n}), 1-\alpha)\right) = 0 \tag{12.17}$$

for $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\} \in SubSeq(h_{2,0})$.

By Lemma A1(a), for $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\} \in SubSeq(h_{2,0})$, we have

$$\nu_{a_n,F_{a_n}}(\theta_{a_n}, \cdot) \Rightarrow \nu_{h_{2,0}}(\cdot) \text{ as } n \to \infty. \tag{12.18}$$

We now show that for all sequences $\tau_n \to \infty$ as $n \to \infty$, we have

$$\lim_{n\to\infty} P_{F_{a_n}}\left(\sup_{g\in\mathcal{G}, j\leq p} |\nu_{a_n,F_{a_n},j}(\theta_{a_n}, g)| > \tau_{a_n}\right) = 0, \tag{12.19}$$

where $\nu_{a_n,F_{a_n},j}(\theta_{a_n}, g)$ denotes the $j$th element of $\nu_{a_n,F_{a_n}}(\theta_{a_n}, g)$. We show this by noting that (12.18) and the continuous mapping theorem give: $\forall \tau > 0$,

$$\lim_{n\to\infty} P_{F_{a_n}}\left(\sup_{g\in\mathcal{G}, j\leq p} |\nu_{a_n,F_{a_n},j}(\theta_{a_n}, g)| > \tau\right) = P\left(\sup_{g\in\mathcal{G}, j\leq p} |\nu_{h_{2,0},j}(g)| > \tau\right), \tag{12.20}$$

where $\nu_{h_{2,0},j}(g)$ denotes the $j$th element of $\nu_{h_{2,0}}(g)$. In addition, the sample paths of $\nu_{h_{2,0},j}(\cdot)$ are bounded a.s., which yields $1\left(\sup_{g\in\mathcal{G}, j\leq p} |\nu_{h_{2,0},j}(g)| > \tau\right) \to 0$ as $\tau \to \infty$ a.s.

Hence, by the bounded convergence theorem,

$$\lim_{\tau\to\infty} P\left(\sup_{g\in\mathcal{G},j\le p} |\nu_{h_{2,0},j}(g)| > \tau\right) = 0. \tag{12.21}$$

Equations (12.20) and (12.21) imply (12.19).

Next, we have

$$\begin{aligned}\xi_{a_n}(\theta_{a_n},g) &= \kappa_{a_n}^{-1}\left(\overline{D}_{a_n}^{-1/2}(\theta_{a_n},g)D_{F_{a_n}}^{1/2}(\theta_{a_n})\right) a_n^{1/2}D_{F_{a_n}}^{-1/2}(\theta_{a_n})\overline{m}_{a_n}(\theta_{a_n},g) \\ &= \kappa_{a_n}^{-1} Diag^{-1/2}(\overline{h}_{2,a_n,F_{a_n}}(\theta_{a_n},g))\left(\nu_{a_n,F_{a_n}}(\theta_{a_n},g)+h_{1,a_n,F_{a_n}}(\theta_{a_n},g)\right)\end{aligned} \tag{12.22}$$

where the second equality holds by the definitions of $\overline{h}_{2,a_n,F_{a_n}}(\theta_{a_n},g)$, $\nu_{a_n,F_{a_n}}(\theta_{a_n},g)$, and $h_{1,a_n,F_{a_n}}(\theta_{a_n},g)$ in (5.2) and $\overline{D}_n(\theta,g) = Diag(\overline{\Sigma}_n(\theta,g))$.

Consider constants $\{\tau_n : n\ge 1\}$ such that $\tau_n\to\infty$ and $\tau_n/\kappa_n\to 0$ as $n\to\infty$. We have

$$\begin{aligned}
&P_{F_{a_n}}\left(c(\varphi_{a_n}(\theta_{a_n}),\widehat{h}_{2,a_n}(\theta_{a_n}),1-\alpha) < c(h_{1,a_n,F_{a_n}}(\theta_{a_n}),\widehat{h}_{2,a_n}(\theta_{a_n}),1-\alpha)\right) \\
&\le P_{F_{a_n}}\left(\varphi_{a_n,j}(\theta_{a_n},g) > h_{1,a_n,F_{a_n},j}(\theta_{a_n},g) \text{ for some } j\le p, \text{ some } g\in\mathcal{G}\right) \\
&\le P_{F_{a_n}}\left(\begin{array}{l}\xi_{a_n,j}(\theta_{a_n},g) > 1 \ \&\ h_{1,a_n,F_{a_n},j}(\theta_{a_n},g) < B_{a_n} \\ \text{for some } j\le p, \text{ some } g\in\mathcal{G}\end{array}\right) \\
&\le P_{F_{a_n}}\left(\begin{array}{l}[\overline{h}_{2,a_n,F_{a_n},j}^{-1/2}(\theta_{a_n},g)\nu_{a_n,F_{a_n},j}(\theta_{a_n},g) + \overline{h}_{2,a_n,F_{a_n},j}^{-1/2}(\theta_{a_n},g)h_{1,a_n,F_{a_n},j}(\theta_{a_n},g)] > \kappa_{a_n} \\ \&\ h_{1,a_n,F_{a_n},j}(\theta_{a_n},g) < B_{a_n} \text{ for some } j\le p, \text{ some } g\in\mathcal{G}\end{array}\right) \\
&\le P_{F_{a_n}}\left(\begin{array}{l}[\tau_{a_n} + \overline{h}_{2,a_n,F_{a_n},j}^{-1/2}(\theta_{a_n},g)h_{1,a_n,F_{a_n},j}(\theta_{a_n},g)] > \kappa_{a_n} \ \& \\ h_{1,a_n,F_{a_n},j}(\theta_{a_n},g) < B_{a_n} \text{ for some } j\le p, \text{ some } g\in\mathcal{G}\end{array}\right) \\
&\quad + P_{F_{a_n}}\left(\sup_{g\in\mathcal{G},j\le p} |\overline{h}_{2,a_n,F_{a_n},j}^{-1/2}(\theta_{a_n},g)\nu_{a_n,F_{a_n},j}(\theta_{a_n},g)| > \tau_{a_n}\right) \\
&\le P_{F_{a_n}}\left(\begin{array}{l}\overline{h}_{2,a_n,F_{a_n},j}^{-1/2}(\theta_{a_n},g)h_{1,a_n,F_{a_n},j}(\theta_{a_n},g) > \kappa_{a_n}-\tau_{a_n} \ \& \\ \overline{h}_{2,a_n,F_{a_n},j}^{-1/2}(\theta_{a_n},g)h_{1,a_n,F_{a_n},j}(\theta_{a_n},g) < \varepsilon^{-1/2}(1+o_p(1))B_{a_n} \\ \text{for some } j\le p, \text{ some } g\in\mathcal{G}\end{array}\right) + o(1) \\
&= o(1),
\end{aligned} \tag{12.23}$$

where the first inequality holds because $c_0(h,1-\alpha+\eta)$ and $c(h,1-\alpha)$ are non-increasing in the first $p$ elements of $h_1$ by Assumption S1(b), the second inequality holds because

$(\theta_{a_n}, F_{a_n}) \in \mathcal{F}$ implies that $h_{1,a_n,F_{a_n},j}(\theta_{a_n}, g) \geq 0\ \forall j \leq p, \forall g \in \mathcal{G}$ and Assumption GMS1(a) implies that (i) $\varphi_{a_n,j}(\theta_{a_n}, g) = 0 \leq h_{1,a_n,F_{a_n},j}(\theta_{a_n}, g)$ whenever $\xi_{a_n,j}(\theta_{a_n}, g) \leq 1$ and (ii) $\varphi_{a_n,j}(\theta_{a_n}, g) \leq B_{a_n}$ a.s. $\forall j \leq p, \forall g \in \mathcal{G}$, the third inequality holds by (12.22), the fourth inequality holds because $P(A) \leq P(A \cap B) + P(B^c)$, the last inequality holds because (i) $\overline{h}_{2,a_n,F_{a_n},j}^{-1/2}(\theta_{a_n}, g) \leq \varepsilon^{-1/2} h_{2,0,j}^{-1/2}(1_k, 1_k)(1 + o_p(1)) = \varepsilon^{-1/2}(1 + o_p(1))$ by Lemma A1(b) and (5.2) and (ii) the second summand on the left-hand side of the last inequality is $o(1)$ by (12.19) with $\tau_{a_n}$ replaced by $\varepsilon^{1/2}\tau_{a_n}/2$ using (i), and the equality holds because $(\kappa_{a_n} - \tau_{a_n}) - \varepsilon^{-1/2}(1 + o_p(1))B_{a_n} = \kappa_{a_n}(1 - \tau_{a_n}/\kappa_{a_n} - \varepsilon^{-1/2}(1 + o_p(1))B_{a_n}/\kappa_{a_n}) = \kappa_{a_n}(1 + o_p(1))$ using Assumption GMS1(b) and $\kappa_{a_n} \to \infty$ as $n \to \infty$.

Hence, (12.17) holds and the Lemma is proved. $\square$

**Proof of Lemma A4.** The result of the Lemma is equivalent to

$$\lim_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}: \\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} P_F\Big(c_0(h_{1,n,F}(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha+\eta) < c_0(h_{1,n,F}(\theta), h_{2,F}(\theta), 1-\alpha) - \varepsilon^*\Big) = 0, \tag{12.24}$$

where $\varepsilon^* = \eta - \delta > 0$. By considering a sequence $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ that is within $\delta_n \to 0$ of the supremum in (12.24) for all $n \geq 1$, it suffices to show that

$$\lim_{n\to\infty} P_{F_n}\Big(c_0(h_{1,n,F_n}(\theta_n), \widehat{h}_{2,n}(\theta_n), 1-\alpha+\eta) < c_0(h_{1,n,F_n}(\theta_n), h_{2,F_n}(\theta_n), 1-\alpha) - \varepsilon^*\Big) = 0. \tag{12.25}$$

Given any subsequence $\{u_n\}$ of $\{n\}$, there exists a subsubsequence $\{a_n\}$ such that $d(h_{2,F_{a_n}}(\theta_{a_n}), h_{2,0}) \to 0$ as $n \to \infty$ for some $h_{2,0} \in \mathcal{H}_{2,cpt}$ because $h_{2,F_n}(\theta_n) \in \mathcal{H}_{2,cpt}$. Hence, it suffices to show that (12.25) holds with $a_n$ in place of $n$.

The condition $d(h_{2,F_{a_n}}(\theta_{a_n}), h_{2,0}) \to 0$ and $(\theta_n, F_n) \in \mathcal{F}$ for all $n \geq 1$ imply that $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\} \in SubSeq(h_{2,0})$. Hence, by Lemma A1(b), $d(\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n}), h_{2,0}) \to_p 0$ as $n \to \infty$. Furthermore,

$$\begin{aligned} &\widehat{h}_{2,a_n}(\theta_{a_n}, g, g^*) \\ &= \widehat{D}_{a_n}^{-1/2}(\theta_{a_n})\widehat{\Sigma}_{a_n}(\theta_{a_n}, g, g^*)\widehat{D}_{a_n}^{-1/2}(\theta_{a_n}) \\ &= Diag(\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n}, 1_k))^{-1/2}\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n}, g, g^*)Diag(\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n}, 1_k))^{-1/2}. \end{aligned} \tag{12.26}$$

Hence, $d(\widehat{h}_{2,a_n}(\theta_{a_n}), h_{2,0}) \to_p 0$ as $n \to \infty$. Given this, using the almost sure representation theorem as above, we can construct $\{\tilde{h}_{2,a_n}(g, g^*) : g, g^* \in \mathcal{G}\}$ such that $d(\tilde{h}_{2,a_n}, h_{2,0}) \to 0$ as $n \to \infty$ a.s. and $\tilde{h}_{2,a_n}$ and $\widehat{h}_{2,a_n}(\theta_{a_n})$ have the same distribution under $(\theta_{a_n}, F_{a_n})$ for all $n \geq 1$.

For fixed $\omega$ in the underlying probability space such that $d(\tilde{h}_{2,a_n}(\cdot,\cdot)(\omega), h_{2,0}) \to 0$ as $n \to \infty$, Lemma A5 with $h_{2,n} = \tilde{h}_{2,a_n}(\omega)$ $(= \tilde{h}_{2,a_n}(\cdot,\cdot)(\omega))$, $h^*_{2,n} = h_{2,F_{a_n}}(\theta_{a_n})$, $h_{2,0} = h_{2,0}$, and $\eta_1 = \eta$ gives: for all $\delta > 0$,

$$\liminf_{n\to\infty} \Big[ c_0(h_{1,a_n,F_{a_n}}(\theta_{a_n}), \tilde{h}_{2,a_n}(\omega), 1-\alpha+\eta) + \delta \\ -c_0(h_{1,a_n,F_{a_n}}(\theta_{a_n}), h_{2,F_{a_n}}(\theta_{a_n}), 1-\alpha) \Big] \geq 0. \tag{12.27}$$

Equation (12.27) holds a.s. This implies that (12.25) holds with $a_n$ in place of $n$ because (i) $\tilde{h}_{2,a_n}$ and $\widehat{h}_{2,a_n}(\theta_{a_n})$ have the same distribution for all $n \geq 1$ and (ii) for any sequence of sets $\{A_n : n \geq 1\}$, $P(A_n\ ev.)$ $(= P(\cup_{m=1}^{\infty} \cap_{k=m}^{\infty} A_k)) = 1$ (where ev. abbreviates eventually) implies that $P(A_n) \to 1$ as $n \to \infty$. $\square$

**Proof of Lemma A5.** Below we show that for $\{h_{2,n}\}$ and $\{h^*_{2,n}\}$ as in the statement of the Lemma, for all constants $x_{h_1,h^*_{2,n}} \in R$ that may depend on $h_1 \in \mathcal{H}_1$ and $h^*_{2,n}$, and all $\delta > 0$,

$$\limsup_{n\to\infty} \sup_{h_1 \in \mathcal{H}_1} \Big[ P(T(h_1, h_{2,n}) \leq x_{h_1,h^*_{2,n}}) - P(T(h_1, h^*_{2,n}) \leq x_{h_1,h^*_{2,n}} + \delta) \Big] \leq 0. \tag{12.28}$$

Note that this result is similar to those of Theorem 1.

We use (12.28) to obtain: for all $\delta > 0$ and $\eta_1 > 0$,

$$\begin{aligned} &\limsup_{n\to\infty} \sup_{h_1 \in \mathcal{H}_1} P(T(h_1, h_{2,n}) \leq c_0(h_1, h^*_{2,n}, 1-\alpha) - \delta) \\ &\leq \limsup_{n\to\infty} \sup_{h_1 \in \mathcal{H}_1} \big[ P(T(h_1, h_{2,n}) \leq c_0(h_1, h^*_{2,n}, 1-\alpha) - \delta) \\ &\qquad\qquad -P(T(h_1, h^*_{2,n}) \leq c_0(h_1, h^*_{2,n}, 1-\alpha) - \delta/2) \big] \\ &\quad + \limsup_{n\to\infty} \sup_{h_1 \in \mathcal{H}_1} P(T(h_1, h^*_{2,n}) \leq c_0(h_1, h^*_{2,n}, 1-\alpha) - \delta/2) \\ &\leq 0 + 1 - \alpha \\ &< 1 - \alpha + \eta_1, \end{aligned} \tag{12.29}$$

where the second inequality holds by (12.28) with $\delta/2$ in place of $\delta$ and $x_{h_1,h^*_{2,n}} =$

$c_0(h_1, h_{2,n}^*, 1-\alpha) - \delta$ and by the definition of the $1-\alpha$ quantile of $T(h_1, h_{2,n}^*)$.

We now use (12.29) to show by contradiction that the result of the Lemma holds. Suppose the result of the Lemma does not hold. Then, there exist constants $\delta > 0$ and $\varepsilon^* > 0$, a subsequence $\{a_n : n \geq 1\}$, and a sequence $\{h_{1,a_n} \in \mathcal{H}_1 : n \geq 1\}$ such that

$$\lim_{n\to\infty} \left[c_0(h_{1,a_n}, h_{2,a_n}, 1-\alpha+\eta_1) + \delta - c_0(h_{1,a_n}, h_{2,a_n}^*, 1-\alpha)\right] \leq -\varepsilon^* < 0. \tag{12.30}$$

Using this and (12.29), we have

$$\begin{aligned}
&\limsup_{n\to\infty} P(T(h_{1,a_n}, h_{2,a_n}) \leq c_0(h_{1,a_n}, h_{2,a_n}, 1-\alpha+\eta_1) + \delta) \\
&\leq \limsup_{n\to\infty} P(T(h_{1,a_n}, h_{2,a_n}) \leq c_0(h_{1,a_n}, h_{2,a_n}^*, 1-\alpha) - \varepsilon^*/2) \\
&\leq \limsup_{n\to\infty} \sup_{h_1\in\mathcal{H}_1} P(T(h_1, h_{2,a_n}) \leq c_0(h_1, h_{2,a_n}^*, 1-\alpha) - \varepsilon^*/2) \\
&< 1-\alpha+\eta_1,
\end{aligned} \tag{12.31}$$

where the first inequality holds by (12.30) and the last inequality holds by (12.29) with $\varepsilon^*/2$ in place of $\delta$.

Equation (12.31) is a contradiction to (12.30) because the left-hand side quantity in (12.31) (without the $\limsup_{n\to\infty}$) is greater than or equal to $1-\alpha+\eta_1$ for all $n \geq 1$ by the definition of the $1-\alpha+\eta_1$ quantile $c_0(h_{1,a_n}, h_{2,a_n}, 1-\alpha+\eta_1)$ of $T(h_{1,a_n}, h_{2,a_n})$. This completes the proof of the Lemma except for establishing (12.28).

To establish (12.28), we write

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{h_1\in\mathcal{H}_1} \left[P(T(h_1, h_{2,n}) \leq x_{h_1,h_{2,n}^*}) - P(T(h_1, h_{2,n}^*) \leq x_{h_1,h_{2,n}^*} + \delta)\right] \qquad (12.32)\\
&\leq \limsup_{n\to\infty} \sup_{h_1\in\mathcal{H}_1} \left[P(T(h_1, h_{2,n}) \leq x_{h_1,h_{2,n}^*}) - P(T(h_1, h_{2,0}) \leq x_{h_1,h_{2,n}^*} + \delta/2)\right] \\
&+\limsup_{n\to\infty} \sup_{h_1\in\mathcal{H}_1} \left[P(T(h_1, h_{2,0}) \leq x_{h_1,h_{2,n}^*} + \delta/2) - P(T(h_1, h_{2,n}^*) \leq x_{h_1,h_{2,n}^*} + \delta)\right].
\end{aligned}$$

The first summand on the right-hand side of (12.32) is less than or equal to 0 by the same argument as used to prove Theorem 1(a) with $\nu_{a_n,F_{a_n}}(\theta_{a_n}, \cdot)$ replaced by $\nu_{h_{2,a_n}}(\cdot)$ in (12.3), where $\nu_{h_{2,a_n}}(\cdot)$ is defined in (4.2), because $d(h_{2,a_n}, h_{2,0}) \to 0$ as $n \to \infty$ implies that the Gaussian processes $\nu_{h_{2,a_n}}(\cdot) \Rightarrow \nu_{h_{2,0}}(\cdot)$ as $n \to \infty$. This argument uses Assumption S2.

Similarly, the second summand on the right-hand side of (12.32) is less than or equal to 0 by an argument analogous to that for Theorem 1(b). Hence, (12.28) is established, which completes the proof. $\square$

# 13 Supplemental Appendix B

## 13.1 Kolmogorov-Smirnov and Approximate CvM Tests and CS's

In this Appendix, we provide results for Kolmogorov-Smirnov (KS) and approximate CvM (A-CvM) tests and CS's defined in Sections 3.1 and 4.2, respectively. A-CvM tests are Cramér-von Mises-type tests in which the test statistic is an infinite sum that is truncated to include only the first $s_n$ functions $\{g_1, ..., g_{s_n}\}$ or the test statistic is an integral with respect to the measure $Q$ and the integral is approximated by a (possibly weighted) average over the functions $\{g_1, ..., g_{s_n}\}$, which are obtained by simulation or by a quasi-Monte Carlo method. The same functions $\{g_1, ..., g_{s_n}\}$ are used for the test statistic and the critical value. In the case of simulated functions, the probabilistic results given here are for fixed (i.e., non-random) functions $\{g_1, ..., g_{s_n}\}$. If $\{g_1, ..., g_{s_n}\}$ are obtained via i.i.d. draws from $Q$, then the probability results are made conditional on the observed functions $\{g_1, ..., g_{s_n}\}$ for $n \geq 1$.

We show that (i) KS and A-CvM CS's have uniform asymptotic coverage probabilities that are greater than or equal to their nominal level $1-\alpha$, (ii) KS and A-CvM tests have asymptotic power equal to one for all fixed alternatives, and (iii) KS and A-CvM tests have asymptotic power that is arbitrarily close to one for a broad array of $n^{-1/2}$-local alternatives whose localization parameter is arbitrarily large.

We consider a slightly more general KS statistic than that defined in (3.7):

$$T_n(\theta) = \sup_{g \in \mathcal{G}_n} S(n^{1/2}\overline{m}_n(\theta, g), \overline{\Sigma}_n(\theta, g)), \tag{13.1}$$

where $\mathcal{G}_n \subset \mathcal{G}$.

For KS tests and CS's, we make use of the following assumptions.

**Assumption KS.** $\mathcal{G}_n \uparrow \mathcal{G}$ as $n \to \infty$.

Let $\mathcal{W}_{bd}$ denote a subset of $\mathcal{W}$ (the set of $k \times k$ positive definite matrices) containing matrices whose eigenvalues are bounded away from zero and infinity.

**Assumption S2$'$.** $S(m, \Sigma)$ is uniformly continuous in the sense that for all bounded sets $\mathcal{M}$ in $R^k$ and all sets $\mathcal{W}_{bd}$

$$\sup_{\mu \in [0,\infty)^p \times \{0\}^v} \sup_{\substack{m, m_0 \in \mathcal{M}: \\ ||m - m_0|| \leq \delta}} \sup_{\substack{\Sigma, \Sigma_0 \in \mathcal{W}_{bd}: \\ ||\Sigma - \Sigma_0|| \leq \delta}} |S(m + \mu, \Sigma) - S(m_0 + \mu, \Sigma_0)| \to 0 \text{ as } \delta \to 0.$$

The following Lemma shows that Assumption S2$'$ is not restrictive.

**Lemma B1.** *The functions $S_1$, $S_2$, and $S_3$ satisfy Assumption* S2$'$.

The following assumption is a strengthening of Assumptions LA1(b) and LA2.

**Assumption LA2$'$.** (a) For all $B < \infty$, $\sup_{g \in \mathcal{G}: h_1(g) \leq B} ||h_{1,n,F_n}(\theta_n, g) - h_1(g)|| \to 0$ as $n \to \infty$, where $\theta_n, F_n$, and $h_1(g)$ are as in Assumption LA1, and

(b) the $k \times d$ matrix $\Pi_F(\theta, g) = (\partial/\partial\theta')[D_F^{-1/2}(\theta) E_F m(W_i, \theta, g)]$ exists and satisfies: for all sequences $\{\delta_n : n \geq 1\}$ such that $\delta_n \to 0$ as $n \to \infty$,

$$\sup_{||\theta - \theta_0|| \leq \delta_n} \sup_{g \in \mathcal{G}} ||\Pi_{F_n}(\theta, g) - \Pi_{F_0}(\theta, g)|| \to 0 \text{ as } n \to \infty \text{ and } \sup_{g \in \mathcal{G}} ||\Pi_{F_0}(\theta_0, g)|| < \infty,$$

where $\theta_0$, $F_0$, and $F_n$ are as in Assumption LA1.

Assumption LA2$'$(a) only requires uniform convergence of $h_{1,n,F_n}(\theta_n, g)$ to $h_1(g)$ over $\{g \in \mathcal{G} : h_1(g) \leq B\}$ because uniform convergence over $g \in \mathcal{G}$ typically does not hold. Assumption LA2$'$ is not restrictive.

For A-CvM tests and CS's, we use Assumptions S2$'$, LA2$'$, and the following assumptions, which hold automatically in the case of an approximate test statistic that is a truncated sum with $s_n \to \infty$.

**Assumption A1.** The functions $\{g_1, ..., g_{s_n}\}$ for $n \geq 1$ are fixed (i.e., non-random) and $s_n \to \infty$ as $n \to \infty$.

**Assumption A2.** The functions $\{g_1, g_2, ...\}$ satisfy:

$$\sum_{\ell=1}^{s_n} w_{Q,n}(\ell) S(m^*(g_\ell), h_{2,F_0}(\theta_*, g_\ell) + \varepsilon I_k) \to \int S(m^*(g), h_{2,F_0}(\theta_*, g) + \varepsilon I_k) dQ(g) \text{ as } n \to \infty,$$

where $m^*(g) = (m_1^*(g), ..., m_k^*(g))'$, $m_j^*(g) = E_{F_0} m_j(W_i, \theta_*) g_j(X_i)/\sigma_{F_0,j}(\theta_*)$, $\theta_*$ and $F_0$ are defined as in Assumption FA, $w_{Q,n}(\ell) = Q(\{g_\ell\})$ in the case of an approximate test statistic that is truncated sum, $w_{Q,n}(\ell) = n^{-1}$ in the case of an approximate test statistic that is a simulated integral, and $w_{Q,n}(\ell)$ is a suitable weight when a test statistic is approximated by a quasi-Monte Carlo method.

**Assumption A3.** The functions $\{g_1, g_2, ...\}$ satisfy: for some sequence of constants $\{B_c^* < \infty : c = 1, 2, ...\}$ such that $B_c^* \to \infty$ as $c \to \infty$,

$$\sum_{\ell=1}^{s_n} w_{Q,n}(\ell) 1(h_1(g_\ell) < B_c^*) S(\Pi_0(g_\ell)\lambda_0, h_2(g_\ell) + \varepsilon I_k)$$
$$\to \int 1(h_1(g) < B_c^*) S(\Pi_0(g)\lambda_0, h_2(g) + \varepsilon I_k) dQ(g) \text{ as } n \to \infty,$$

where $\Pi_0(g) = \Pi_{F_0}(\theta_0, g)$, $h_2(g) = h_{2,F_0}(\theta_0, g)$, and $\theta_0$ and $F_0$ are defined as in Assumption LA1.

Assumptions A1-A3 are not restrictive because (i) they hold automatically if the approximate test statistic is a truncated sum and (ii) if the approximate test statistic is a simulated integral and $\{g_1, g_2, ...\}$ are i.i.d. with distribution $Q$ and $s_n \to \infty$ as $n \to \infty$, then they hold conditional on $\{g_1, g_2, ...\}$ with probability one.

The following result establishes that nominal $1-\alpha$ KS and A-CvM CS's have uniform asymptotic coverage probability greater than or equal to $1 - \alpha$.

**Theorem B1.** *Suppose Assumptions* M, S1, *and* S2$'$ *hold and Assumption* GMS1 *holds when considering* GMS CS's. *Then, for every compact subset* $\mathcal{H}_{2,cpt}$ *of* $\mathcal{H}_2$, KS-GMS, KS-PA, A-CvM-GMS, *and* A-CvM-PA *confidence sets* $CS_n$ *satisfy*

$$\liminf_{n\to\infty} \inf_{\substack{(\theta,F)\in\mathcal{F}: \\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} P_F(\theta \in CS_n) \geq 1 - \alpha.$$

**Comments. 1.** Assumptions KS and A1 are not needed in Theorem B1.

**2.** Theorem B1 is an analogue of Theorem 2(a) for CS's based on KS and A-CvM statistics. It is proved by making adjustments to the proof of Theorem 2(a). An analogue of Theorem 2(b) is not given here because the proof of Theorem 2(b) does not go through with KS or A-CvM test statistics. The proof of Theorem 2(b) utilizes the bounded convergence theorem which applies only if the test statistic is an integral with respect to some measure $Q$. The continuous mapping theorem cannot be applied because the convergence of $h_{1,n,F_n}(\theta_n, g)$ to $h_{1,\infty,F_0}(\theta_0, g)$ is not uniform over $g \in \mathcal{G}$ for many sequences $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$, where $(\theta_n, F_n) \to (\theta_0, F_0)$.

The next result shows that KS and A-CvM tests have asymptotic power equal to one against all fixed alternatives. This implies that any parameter value outside the identified set is included in a KS or A-CvM CS with probability that goes to zero as $n \to \infty$, see the Comment to Theorem 3.

**Theorem B2.** *Suppose Assumptions* FA, CI, Q, S1, S3, *and* S4 *hold, Assumption* KS *holds when considering the* KS *test, and Assumptions* A1 and A2 *hold when considering* A-CvM *tests. Then, the* KS-GMS and KS-PA *tests satisfy the results of Theorem* 3 *concerning power under fixed alternatives. In addition,* A-CvM-GMS *and* A-CvM-PA *tests, respectively, satisfy*

(a) $\lim_{n\to\infty} P_{F_0}(\overline{T}_{n,s_n}(\theta_*) > c_{s_n}(\varphi_n(\theta_*), \widehat{h}_{2,n}(\theta_*), 1-\alpha)) = 1$ *and*

(b) $\lim_{n\to\infty} P_{F_0}(\overline{T}_{n,s_n}(\theta_*) > c_{s_n}(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_*), 1-\alpha)) = 1$.

The following result is for $n^{-1/2}$-local alternatives.

**Theorem B3.** *Suppose Assumptions* M, S1-S4, S2$'$, LA1, *and* LA2$'$ *hold, Assumptions* KS *and* LA3 *hold when considering the* KS *test, and Assumptions* A1, A3, *and* LA3$'$ *hold when considering* A-CvM *tests. Let* $\theta_{n,*} = \theta_{n,*}(\beta) = \theta_n + \beta\lambda_0 n^{-1/2}(1+o(1))$ *be as in Assumption* LA1(a) *with* $\lambda = \beta\lambda_0$ *for some* $\beta > 0$ *and* $\lambda_0 \in R^{d_\theta}$. *Then, under* $n^{-1/2}$*-local alternatives, the* A-CvM-GMS *and* A-CvM-PA *tests, respectively, satisfy*

(a) $\lim_{\beta\to\infty}\lim_{n\to\infty} P_{F_n}(\overline{T}_{n,s_n}(\theta_{n,*}(\beta)) > c_{s_n}(\varphi_n(\theta_{n,*}(\beta)), \widehat{h}_{2,n}(\theta_{n,*}(\beta)), 1-\alpha)) = 1$ *provided Assumption* GMS1 *also holds,*

(b) $\lim_{\beta\to\infty}\lim_{n\to\infty} P_{Fn}(\overline{T}_{n,s_n}(\theta_{n,*}(\beta)) > c_{s_n}(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_{n,*}(\beta)), 1-\alpha)) = 1$, *and*

(c) KS-GMS *and* KS-PA *tests satisfy parts* (a) *and* (b), *respectively, with* $\overline{T}_{n,s_n}(\theta_{n,*}(\beta))$ *replaced by* $T_n(\theta_{n,*}(\beta))$ *and with the subscript* $s_n$ *on* $c_{s_n}(\cdot,\cdot,\cdot)$ *deleted.*

**Comment.** Theorem B3 shows that KS and A-CvM tests have power arbitrarily close to one for the same $n^{-1/2}$-local alternatives as Cramér-von Mises tests that are based on integrals with respect to a probability measure $Q$.

## 13.2 Instruments and Weight Functions

In this section we provide three additional examples of instruments $\mathcal{G}$ and weight functions $Q$ that satisfy Assumptions CI, M, F(e), and Q. We also specify non-data-dependent methods for transforming a regressor to lie in $[0,1]$.

If $x \in R$ is known to lie in an open, closed, or half-open interval denoted by $\lfloor c, d \rfloor$, where $-\infty \leq c \leq d \leq \infty$, then one can transform $x$ into $[0,1]$ via

$$\begin{array}{llll} t(x) = \frac{x-c}{d-c} & \text{if } c > -\infty \ \& \ d < \infty, & t(x) = \frac{e^x}{1+e^x} & \text{if } c = -\infty \ \& \ d = \infty, \\ t(x) = \frac{e^{x-c}-1}{1+e^{x-c}} & \text{if } c > -\infty \ \& \ d = \infty, & t(x) = \frac{2e^{x-d}}{1+e^{x-d}} & \text{if } c = -\infty \ \& \ d < \infty. \end{array} \tag{13.2}$$

Alternatively, a vector $X_i$ can be transformed first to have sample mean equal to zero and sample variance matrix equal to $I_{d_x}$ (by multiplication by the inverse of the upper-triangular Cholesky decomposition of the sample covariance matrix of $X_i$). Then, it can be transformed to lie in $[0,1]^{d_x}$ by applying the standard normal distribution function $\Phi(\cdot)$ element by element. This method is employed in Section 10.3.

**Example 3. (B-splines).** A collection of B-splines provides a set $\mathcal{G}$ that satisfies Assumptions CI and M for those $(\theta, F)$ for which $E_F(m_j(W_i,\theta)|X_i = x)$ is a continuous function of $x$ for all $j \leq k$. The regressors are transformed to lie in $[0,1]^{d_x}$. We consider normalized cubic B-splines with equally-spaced knots on $[0,1]^{d_x}$. (B-splines of other orders also could be considered.) The class of normalized cubic B-splines is a countable set defined by

$$\begin{aligned} \mathcal{G}_{B-spline} &= \{g(x) : g(x) = B_C(x) \cdot 1_k \text{ for } C \in \mathcal{C}_{B-spline}\}, \text{ where} \\ \mathcal{C}_{B-spline} &= \Big\{C^*_{a,r} = \times_{u=1}^{d_x} \left[((a_u - 1)/(2r), (a_u+3)/(2r)] \cap [0,1]\right] \in [0,1]^{d_x} : a = (a_1, ..., a_{d_x})' \\ &\qquad a_u \in \{-2,-1,...,2r\} \text{ for } u = 1,...,d_x \text{ and } r = r_0, r_0+1, ...\Big\} \text{ and} \\ B_{C^*_{a,r}}(x) &= 1(x \in C^*_{a,r}) \\ &\quad \times \prod_{u=1}^{d_x} \begin{cases} y_u^3/6 & \text{for } x_u \in ((a_u-1)/(2r), a_u/(2r)] \\ (-3y_u^3 + 12y_u^2 - 12y_u + 4)/6 & \text{for } x_u \in (a_u/(2r), (a_u+1)/(2r)] \\ (-3z_u^3 + 12z_u^2 - 12z_u + 4)/6 & \text{for } x_u \in ((a_u+1)/(2r), (a_u+2)/(2r)] \\ z_u^3/6 & \text{for } x_u \in ((a_u+2)/(2r), (a_u+3)/(2r)] \\ 0 & \text{otherwise,} \end{cases} \\ &\quad x = (x_1,...,x_{d_x})', \ y_u = 2rx_u - (a_u - 1), \text{ and } z_u = 4 - y_u \text{ for } u = 1,...,d_x, \end{aligned} \tag{13.3}$$

for some positive integer $r_0$, see Schumaker (2007, p. 136). If $d_x = 1$, a B-spline in $\mathcal{G}_{B-spline}$ has finite support given by the union of four consecutive subintervals each of length $(2r)^{-1}$. If $d_x \geq 1$, a cubic B-spline in $\mathcal{G}_{B-spline}$ has support on a $d_x$-dimensional hypercube in $[0,1]^{d_x}$ with edges of length $4 \cdot (2r)^{-1}$.

Note that a bounded continuous product kernel with bounded support could be used in place of B-splines in Example 3.

**Weight Function Q for $\mathcal{G}_{B-spline}$.** There is a one-to-one mapping $\Pi_{B-spline} : \mathcal{G}_{B-spline} \rightarrow AR^*$, where $AR^*$ is defined as $AR$ is defined in Section 3.4 but with $\{-2,-1,...,2r\}^{d_x}$ in place of $\{1,...,2r\}^{d_x}$. We take $Q = \Pi^{-1}_{B-spline} Q_{AR^*}$, where $Q_{AR^*}$ is a probability measure on $AR^*$. For example, the uniform distribution on $a \in \{-2,-1,...,2r\}^{d_x}$ conditional on $r$ and some discrete mass function $\{w(r) : r = r_0, r_0+1, ...\}$ on $r$ gives the test statistic:

$$T_n(\theta) = \sum_{r=r_0}^{\infty} w(r) \sum_{a \in \{-2,-1,...,2r\}^{d_x}} (2r+3)^{-d_x} S(n^{1/2}\overline{m}_n(\theta, g_{a,r}), \overline{\Sigma}_n(\theta, g_{a,r})), \qquad (13.4)$$

where $g_{a,r}(x) = B_{C^*_{a,r}}(x) \cdot 1_k$ for $C^*_{a,r} \in \mathcal{C}_{B-spline}$

**Example 4 (Data-dependent Boxes).** Next, we consider a class of functions $\mathcal{G}_{box,dd}$ that is designed to be applied with a data-dependent weight function $Q$ defined below. Because this $Q$ only puts positive weight on center-points $x$ that are in the support of $X_i$, it turns out to be necessary to consider boxes with different left and right edge lengths as measured from the "center" point. (See footnote 46 below for an explanation.)

We define

$$\begin{aligned} \mathcal{G}_{box,dd} &= \{g : g(x) = 1(x \in C) \cdot 1_k \text{ for } C \in \mathcal{C}_{box,dd}\}, \text{ where} \qquad (13.5) \\ \mathcal{C}_{box,dd} &= \left\{C_{x,r_1,r_2} = \times_{u=1}^{d_x}(x_u - r_{1,u}, x_u + r_{2,u}] : x \in Supp_{F_{X,0}}(X_i), r_{1,u}, r_{2,u} \in (0, \bar{r}) \ \forall u \leq d_x\right\} \end{aligned}$$

for some $\bar{r} \in (0, \infty]$, $x = (x_1, ..., x_{d_x})'$, $r_1 = (r_{1,1}, ..., r_{1,d_x})'$, $r_2 = (r_{2,1}, ..., r_{2,d_x})'$, and $Supp_{F_{X,0}}(X_i)$ denotes the support of $X_i$ when $F_0$ is the true distribution.

**Data-dependent Q for $\mathcal{G}_{box,dd}$.** There is a one-to-one mapping $\Pi_{box,dd} : \mathcal{G}_{box,dd} \rightarrow \{(x, r_1, r_2) \in Supp_{F_{X,0}}(X_i) \times (0, \bar{r})^{2d_x}\}$. Thus, for any probability measure $Q^*$ on $\{(x, r_1, r_2) \in Supp_{F_{X,0}}(X_i) \times (0, \bar{r})^{2d_x}\}$, $(\Pi_{box,dd})^{-1}Q^*$ is a valid probability measure on $\mathcal{G}_{box,dd}$. In this case, the inverse mapping $(\Pi_{box,dd})^{-1}$ is $(\Pi_{box,dd})^{-1}[x, r_1, r_2] = g_{x,r_1,r_2}(\cdot) = 1(\cdot \in C_{x,r_1,r_2}) \cdot 1_k$. Let

$$\begin{aligned} Q^*_{F_{X,0}} &= F_{X,0} \times Unif\left(\left(\times_{u=1}^{d_x}(0, \sigma_{X,u}\bar{r})\right)^2\right), \text{ where} \\ \sigma^2_{X,u} &= Var_{F_{X,0}}(X_{i,u}) \text{ for } u = 1, ..., d_x \qquad (13.6) \end{aligned}$$

and $F_{X,0}$ denotes the true distribution of $X_i$.[46] The scale factors $\sigma_{X,1}, ..., \sigma_{X,d_x}$ are included here to make $Q^*_{F_{X,0}}$ equivariant to location and scale changes in $X_i$. Of course, $F_{X,0}$ and $\{\sigma^2_{X,u} : u \leq d_x\}$ are unknown, so they need to be replaced by estimators. The distribution $F_{X,0}$ can be estimated by the empirical distribution of $X_i$ based on a subsample of size $b_n$ of $\{X_i : i \leq n\}$, denoted by $\widehat{F}_{X,b_n}(\cdot)$. Here we use the empirical distribution based on a subsample, rather than the whole sample, because the computational costs are large when $b_n = n$ and $n$ is large.[47] The variances $\{\sigma^2_{X,u} : u \leq d_x\}$ can be estimated by the sample variances based on $\{X_i : i \leq n\}$, denoted by $\{\widehat{\sigma}^2_{X,n,u} : u = 1, ..., d_x\}$. In this case, the test statistic is

$$\begin{aligned}
&T_n(\theta) \\
&= \int_{R^{d_x}} \int_{\left(\times_{u=1}^{d_x}(0,\widehat{\sigma}_{X,n,u}\bar{r})\right)^2} S(n^{1/2}\overline{m}_n(\theta, g_{x,r_1,r_2}), \overline{\Sigma}_n(\theta, g_{x,r_1,r_2})) \\
&\qquad\qquad \times \prod_{u=1}^{d_x} (\widehat{\sigma}_{X,n,u}\bar{r})^{-2} dr_1 dr_2 d\widehat{F}_{X,m_n}(x) \qquad (13.7) \\
&= b_n^{-1} \sum_{i=1}^{b_n} \int_{\left(\times_{u=1}^{d_x}(0,\widehat{\sigma}_{X,n,u}\bar{r})\right)^2} S(n^{1/2}\overline{m}_n(\theta, g_{X_i,r_1,r_2}), \overline{\Sigma}_n(\theta, g_{X_i,r_1,r_2})) dr_1 dr_2 \prod_{u=1}^{d_x} (\widehat{\sigma}_{X,n,u}\bar{r})^{-2},
\end{aligned}$$

where $g_{x,r_1,r_2}$ is as above.

When an approximate test statistic $\overline{T}_{n,s_n}(\theta)$ that is a simulated integral is employed, see (3.16) in Section 3.5, it is defined as in (13.7) but with the integral over $(r_1, r_2)$ replaced by an average over $\ell = 1, ..., s_n$, the term $\prod_{u=1}^{d_x} (\widehat{\sigma}_{X,n,u}\bar{r})^{-2}$ deleted, and $g_{X_i,r_1,r_2}$ replaced by $g_{X_i,r_{1,\ell},r_{2,\ell}}$, where $\{(r_{1,\ell}, r_{2,\ell}) : \ell = 1, ..., s_n\}$ are i.i.d. with a $Unif(\times_{u=1}^{d_x}(0, \widehat{\sigma}_{X,n,u}\bar{r}))^2$ distribution. Alternatively, in this case, one can take $b_n = s_n$, delete the integral over $(r_1, r_2)$, delete the term $\prod_{u=1}^{d_x} (\widehat{\sigma}_{X,n,u}\bar{r})^{-2}$, and replace $g_{X_i,r_1,r_2}$ by $g_{X_i,r_{1,i},r_{2,i}}$, where $\{(r_{1,i}, r_{2,i}) : i = 1, ..., s_n\}$ are as above.

---

[46] One might think that a natural data-dependent measure $Q$ is $Q^s = \Pi^{-1}_{box}(F_{X,0} \times Unif((0, \bar{r})^{d_x})$, defined on $\mathcal{G}^s_{box}$, where $\mathcal{G}^s_{box}$ is defined as $\mathcal{G}_{box}$ is defined in (3.13) but with $R$ replaced by $Supp(X_i)$. However, such a $Q$ does not necessarily have support that contains $\mathcal{G}^s_{box}$ and, hence, the resulting test may not have power against all fixed alternatives. See the following paragraph for details. It is for this reason that $\mathcal{G}_{box,dd}$ is defined to contain boxes that are asymmetric about their center points.

The probability distribution $Q^s$ on $\mathcal{G}^s_{box}$, does not necessarily satisfy Assumption Q. To see why, consider a simple example with $d_x = 1$ and $k = 1$. Suppose $X_i$ takes only four values: 0, 1, 2, 3 each with probability 1/4 and $\bar{r} > 1$. Then, for $g_{1,1}(x) = 1(x \in (0, 2]) \in \mathcal{G}^s_{box}$, we have $\mathcal{B}(g_{1,1}, \delta) = \{g_{1,1}\}$. This holds because if $\omega > 0$, $g_{1,1+\omega}(0) = 1$ but $g_{1,1}(0) = 0$; if $\omega < 0$, $g_{1,1+\omega}(2) = 0$ but $g_{1,1}(2) = 1$; if $\omega > 0$, $g_{2,1+\omega}(3) = 1$ but $g_{1,1}(3) = 0$; and if $\omega < 0$, $g_{2,1+\omega}(1) = 0$ but $g_{1,1}(1) = 1$. The set $\{g_{1,1}\}$ has zero $Q^s$ measure. So, $Q^s$ does not satisfy Assumption Q.

[47] Also, it is easier to establish the asymptotic validity of this procedure when $b_n/n \to 0$ as $n \to \infty$.

**Example 5. (Continuous/Discrete Regressors).** The collections $\mathcal{G}_{c\text{-}cube}$ and $\mathcal{G}_{box}$ (defined in the main paper) and $\mathcal{G}_{B-spline}$ and $\mathcal{G}_{box,dd}$ (defined here) can be used with continuous and/or discrete regressors. However, one can design $\mathcal{G}$ to exploit the known support of discrete regressors. Suppose $X_i = (X_{1,i}', X_{2,i}')'$, where $X_{1,i} \in R^{d_{x,1}}$ is a continuous random vector and $X_{2,i} \in R^{d_{x,2}}$ is a discrete random vector that takes values in a countable set $D = \{x_{2,1}, x_{2,2}, ...\}$, where $x_{2,u} \in R^{d_{x,2}}$ for all $u \geq 1$. Define the set $\mathcal{G}_{c/d}$ by

$$\mathcal{G}_{c/d} = \{g : g = g_1 g_2,\ g_1 \in \mathcal{G}_1,\ g_d \in \mathcal{G}_D\}, \tag{13.8}$$

where $x = (x_1', x_2')'$, $g_1$ is an $R^k$-valued function of $x_1$, $g_2$ is an $R$-valued function of $x_2$, $\mathcal{G}_1 = \mathcal{G}_{c\text{-}cube}$, $\mathcal{G}_{box}$, $\mathcal{G}_{B-spline}$, or $\mathcal{G}_{box,dd}$, with $x$ and $d_x$ replaced by $x_1$ and $d_{x,1}$, respectively, and $\mathcal{G}_D = \{g_d : g_d(x_2) = 1_{\{d\}}(x_2)\}$ for $d \in D\}$.

**Weight Function Q for $\mathcal{G}_{c/d}$.** When $\mathcal{G}$ is of the form $\mathcal{G}_{c/d}$, it is natural to take $Q$ to be of the form $Q_1 \times Q_D$, where $Q_1$ is a probability measure on $\mathcal{G}_1$, such as any of those considered above with $x_1$ in place of $x$, and $Q_D$ is a probability measure on $D$. If $D$ is a finite set, then one may take $Q_D$ to be uniform. For example, when $\mathcal{G}_1 = \mathcal{G}_{box}$ and $Q_D$ is uniform, the test statistic is

$$T_n(\theta) = \frac{1}{\#D} \sum_{d \in D} \int_{[0,1]^{d_{x,1}}} \int_{(0,\bar{r})^{d_{x,1}}} S(n^{1/2}\overline{m}_n(\theta, g_{x_1,r} g_d), \overline{\Sigma}_n(\theta, g_{x_1,r} g_d)) \bar{r}^{-d_x} dr dx_1, \tag{13.9}$$

where $\#D$ denotes the number of elements in $D$ and $x_1 \in R^{d_{x,1}}$. When $\mathcal{G}_1 = \mathcal{G}_{c\text{-}cube}$ or $\mathcal{G}_{B-spline},$ $T_n(\theta)$ is a combination of the formulae given above.

The following result establishes Assumptions CI, M, and FA(e) for $\mathcal{G}_{B-spline}$, $\mathcal{G}_{box,dd}$, and $\mathcal{G}_{c/d}$ and Assumption Q for the weight functions $Q$ on these sets.

**Lemma B2.** (a) *For any moment function $m(W_i, \theta)$, Assumptions* CI *and* M *hold with $\mathcal{G} = \mathcal{G}_{B-spline}$ for all $(\theta, F)$ for which $E_F(m_j(W_i, \theta)|X_i = x)$ is a continuous function of $x$ for all $j \leq k$.*

(b) *For any moment function $m(W_i, \theta)$, Assumptions* CI *and* M *hold with $\mathcal{G} = \mathcal{G}_{box,dd}$.*

(c) *For any moment function $m(W_i, \theta)$, Assumptions* CI *and* M *hold with $\mathcal{G} = \mathcal{G}_{c/d}$, where $\mathcal{G}_1 = \mathcal{G}_{c\text{-}cube}$, $\mathcal{G}_{box}$, $\mathcal{G}_{B-spline}$, or $\mathcal{G}_{box,dd}$, with $(x, d_x)$ replaced by $(x_1, d_{x,1})$ and in the case of $\mathcal{G}_1 = \mathcal{G}_{B-spline}$ Assumption* CI *and* M *only hold for $(\theta, F)$ for which $E_F(m_j(W_i, \theta)|X_{i,1} = x_1, X_{2,i} = d)$ is a continuous function of $x_1 \in [0,1]^{d_{x,1}}$ $\forall d \in D$,*

$\forall j \leq k$.

(d) *Assumption* FA(e) *holds for* $\mathcal{G}_{B-spline}$, $\mathcal{G}_{box,dd}$, *and* $\mathcal{G}_{c/d}$.

(e) *Assumption* Q *holds for the weight function* $Q_c = \Pi_{B-spline}^{-1} Q_{AR^*}$ *on* $\mathcal{G}_{B-spline}$, *where* $Q_{AR^*}$ *is uniform on* $a \in \{-2, -1, ..., 2r\}^{d_x}$ *conditional on* $r$ *and* $r$ *has some probability mass function* $\{w(r) : r = r_0, r_0 + 1, ...\}$ *with* $w(r) > 0$ *for all* $r$.

(f) *Assumption* Q *holds for the weight function* $Q_d = (\Pi_{box,dd})^{-1} Q^*_{F_{X,0}}$, where $Q^*_{F_{X,0}} = (F_{X,0} \times Unif((\times_{u=1}^{d_x}(0, \sigma_{X,u}\bar{r}))^2)$ on $\mathcal{G}_{box,dd}$.

(g) *Assumption* Q *holds for the weight function* $Q_e = Q_1 \times Q_D$ *on* $\mathcal{G}_{c/d}$, *where* $Q_1$ *is a probability measure on* $\mathcal{G}_1$ *equal to any of the distributions* $Q$ *on* $\mathcal{G}$ *considered in part* (e), *part* (f), *or in Lemma* 4 *but with* $x_1$ *in place of* $x$, $D$ *is a finite set, and* $Q_D = Unif(D)$.

**Comment.** The uniform distribution that appears in parts (e)-(g) of the Lemma could be replaced by another distribution and the results of the Lemma still hold provided the other distribution has the same support. For example, in part (g), Assumption Q holds when $D$ is a countably infinite set and $Q_D$ is a probability measure whose support is $D$.

## 13.3 Example: Verification of Assumptions LA1-LA3 and LA3′

Here we verify Assumptions LA1-LA3 and LA3′ in a simple example for purposes of illustration. These assumptions are the main assumptions employed with local alternatives.

**Example.** Suppose $W_i = (Y_i, X_i)' \in R^2$ and there is a single moment inequality function $m(W_i, \theta) = Y_i - \theta$ and no moment equalities, i.e., $p = 1$ and $v = 0$. Suppose the true parameters/distributions $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ and the null values $\{\theta_{n,*} \in \Theta, : n \geq 1\}$ satisfy: (i) $\theta_n \to \theta_0$ and $F_n \to F_0$ (under the Kolmogorov metric) for some $(\theta_0, F_0) \in \mathcal{F}$, (ii) $\theta_{n,*} = \theta_n + \lambda n^{-1/2}$ for some $\lambda > 0$, (iii) $Y_i = \theta_n + \mu(X_i)n^{-1/2} + U_i$, (iv) $\mu(x) \geq 0$, $\forall x \in R$, and (v) under all $F$ such that $(\theta, F) \in \mathcal{F}$ for some $\theta \in \Theta$, $(X_i, U_i)$ are i.i.d. with distribution that does not depend on $F$, $X_i$ and $U_i$ are independent, $E_F U_i = 0$, $Var_F(U_i) = 1$, $Var_F(X_i) \in (0, \infty)$, and $E_F|U_i|^{2+\delta} + E_F|\mu(X_i)|^{2+\delta} < \infty$ for some $\delta > 0$, and $\sup_{g \in \mathcal{G}} E_F(1 + \mu^2(X_i))(1 + g^2(X_i)) < \infty$.

We show that in this example Assumptions LA1 and LA2 hold, Assumption LA3 holds if $\lambda$ is sufficiently large, and Assumption LA3′ holds if $\mathcal{G}$ and $Q$ satisfy Assumptions CI and Q, respectively.

By (v), we can write $E_F g(X_i) = Eg(X_i)$ and $E_F \mu(X_i)g(X_i) = E\mu(X_i)g(X_i)$.

Assumption LA1(a) holds by (i) and (ii). Assumption LA1(b) holds by the following calculations:

$$
\begin{aligned}
&n^{1/2} E_{F_n} m(W_i, \theta_n, g) = n^{1/2} E_{F_n} (U_i + \mu(X_i) n^{-1/2}) g(X_i) = h_1(g), \text{ where} \\
&h_1(g) = E\mu(X_i) g(X_i) \in [\,0, \infty) \text{ and} \qquad\qquad (13.10) \\
&\sigma^2_{F_n}(\theta_n) = Var_{F_n}(Y_i) = Var_{F_n}(U_i + \mu(X_i) n^{-1/2}) = 1 + n^{-1} Var_{F_n}(\mu(X_i)) \to 1.
\end{aligned}
$$

To show Assumption LA1(c), we have

$$
\begin{aligned}
E_{F_n} Y_i^2 g(X_i) g^*(X_i) &= E_{F_n} (\theta_n + \mu(X_i) n^{-1/2} + U_i)^2 g(X_i) g^*(X_i) \\
&\to E_{F_0} (\theta_0 + U_i)^2 g(X_i) g^*(X_i) \\
&= E_{F_0} Y_i^2 g(X_i) g^*(X_i) \text{ as } n \to \infty, \qquad (13.11)
\end{aligned}
$$

uniformly over $g, g^* \in \mathcal{G}$, using (i), (iii), and (v). Here we have used $Y_i = \theta_0 + U_i$ under $F_0$. This holds because $F_n \to F_0$ by (ii), which implies that $P_{F_n}(Y_i \leq y) \to P_{F_0}(Y_i \leq y)$ for all continuity points $Y_i$, but direct calculations show that $P_{F_n}(Y_i \leq y) = P(\theta_n + \mu(X_i) n^{-1/2} + U_i \leq y) \to P(\theta_0 + U_i \leq y)$ for all continuity points $y$ of $U_i + \theta_0$ and, hence, $Y_i = \theta_0 + U_i$ under $F_0$.

Next, we write

$$
\begin{aligned}
&E_{F_n} m(W_i, \theta_n, g) m(W_i, \theta_n, g^*) \\
&= E_{F_n} Y_i^2 g(X_i) g^*(X_i) - \theta_n E[E_{F_n}(Y_i | X_i)(g(X_i) + g^*(X_i))] + \theta_n^2 E g(X_i) g^*(X_i) \\
&= E_{F_n} Y_i^2 g(X_i) g^*(X_i) - \theta_n E[(\theta_n + \mu(X_i) n^{-1/2})(g(X_i) + g^*(X_i))] \\
&\quad + \theta_n^2 E g(X_i) g^*(X_i) \\
&= E_{F_0} Y_i^2 g(X_i) g^*(X_i) - \theta_0^2 E g(X_i) - \theta_0^2 E g^*(X_i) + \theta_0^2 E g(X_i) g^*(X_i) + o(1) \\
&= E_{F_0} m(W_i, \theta_0, g) m(W_i, \theta_0, g^*) + o(1), \qquad (13.12)
\end{aligned}
$$

where $o(1)$ holds uniformly over $g, g^* \in \mathcal{G}$, using (13.11), (i), (iii), and (v). In addition, $E_{F_n} m(W_i, \theta_n, g) = o(1)$ and $E_{F_0} m(W_i, \theta_0, g) = o(1)$ uniformly over $g \in \mathcal{G}$ by (13.10) and (v). Hence, the first part of Assumption LA1(c) holds. The second part of Assumption LA1(c) holds by the same argument with $\theta_{n,*}$ in place of $\theta_n$.

Assumption LA1(d) holds because $Var_{F_n}(m_j(W_i, \theta_{n,*})) = Var_{F_n}(m_j(W_i, \theta_n)) > 0$. Assumption LA1(e) holds using (v) and the above expression for $\sigma^2_{F_n}(\theta_n)$.

Assumption LA2 holds because $\Pi_F(\theta, g)$ does not depend on $(\theta, F)$ by the following calculations and (v): $\forall F$ such that $(\theta, F) \in \mathcal{F}$ and $\forall g \in \mathcal{G}$,

$$\begin{aligned}\Pi_F(\theta, g) &= (\partial/\partial\theta)[D_F^{-1/2}(\theta)E_F m(W_i, \theta, g)] \\ &= \sigma_F^{-1}(\theta)(\partial/\partial\theta)E_F(Y_i - \theta)g(X_i)] = -\sigma_F^{-1}(\theta)Eg(X_i),\end{aligned} \tag{13.13}$$

where the second equality holds because $D_F(\theta) = \sigma_F^2(\theta) = Var_F(Y_i)$ does not depend on $\theta$.

We have: $\Pi_0(g) = \Pi_{F_0}(\theta_0, g) = -Eg(X_i)$ by (13.13) and $\sigma_{F_0}^2(\theta_0) = 1$. Hence, in Assumption LA3, $h_1(g)+\Pi_0(g)\lambda = E\mu(X_i)g(X_i)-Eg(X_i)\lambda$, which is negative whenever $\lambda > E\mu(X_i)g(X_i)/Eg(X_i)$. Hence, if the null value $\theta_{n,*}$ deviates from the true value $\theta_n$ by enough (i.e., if $n^{1/2}(\theta_{n,*}-\theta_n) = \lambda$ is large enough), then the null hypothesis is violated for all $n$ and Assumption LA3 holds.

Next, we show that Assumption LA3$'$ holds provided Assumptions CI and Q hold. We have: (a) $\Pi_0(g) = -Eg(X_i)$, (b) $h_1(g) < \infty$ $\forall g \in \mathcal{G}$ by (13.10) using (v), and (c) $\lambda_0 = \lambda/\beta > 0$ because $\lambda > 0$ by (ii) and $\beta > 0$ by definition. Hence, the condition of Assumption LA3$'$ reduces to

$$Q(\{g \in \mathcal{G} : Eg(X_i) > 0\}) > 0. \tag{13.14}$$

Suppose $Eg^*(X_i) > 0$ for some $g^* \in \mathcal{G}$. (This is a very weak requirement on $\mathcal{G}$ and is implied by Assumption CI, see below.) Let $\delta_1 = Eg^*(X_i) > 0$. Then, using the metric $\rho_X$ defined in Section 6, for any $g \in \mathcal{G}$ with $\rho_X(g, g^*) < \delta_1$, we have $Eg(X_i) > 0$ because otherwise $g(X_i) = 0$ a.s. and $\delta_1 > \rho_X(g, g^*) = (Eg^*(X_i)^2)^{1/2} \geq Eg^*(X_i) = \delta_1$, which is a contradiction. Thus, $Eg(X_i) > 0$ for all $g \in \mathcal{B}_{\rho_X}(g^*, \delta_1)$, where $\mathcal{B}_{\rho_X}(g^*, \delta_1)$ is the open $\rho_X$-ball in $\mathcal{G}$ centered at $g^*$ with radius $\delta_1$. By Assumption Q, $Q(\mathcal{B}_{\rho_X}(g^*, \delta_1)) > 0$. Hence, (13.14) holds and Assumption LA3$'$ is verified.

Lastly, we show that Assumption CI implies that $Eg^*(X_i) > 0$ for some $g^* \in \mathcal{G}$. For all $\theta > \theta_0$, we have

$$\begin{aligned}\mathcal{X}_{F_0}(\theta) &= \{x \in R : E_{F_0}(m_j(W_i, \theta)\,|X_i = x) < 0\} \\ &= \{x \in R : \theta_0 - \theta < 0\} = R,\end{aligned} \tag{13.15}$$

where the second equality holds because $Y_i = \theta_0 + U_i$ under $F_0$, and so, $E_{F_0}(m_j(W_i, \theta)\,|X_i = x) = E_{F_0}(Y_i - \theta | X_i = x) = \theta_0 - \theta$.

By (13.15), $P_{F_0}(X_i \in \mathcal{X}_{F_0}(\theta)) = P_{F_0}(X_i \in R) = 1 > 0$. Hence, by Assumption CI, there exists $g^* \in \mathcal{G}$ such that $E_{F_0} m(W_i, \theta) g^*(X_i) = E(\theta_0 - \theta) g^*(X_i) < 0$ for $\theta > \theta_0$. That is, $Eg^*(X_i) > 0$.

## 13.4 Uniformity Issues with Infinite-Dimensional Nuisance Parameters

This section illustrates one of the subtleties that arises when considering the uniform asymptotic behavior of a test or CS in a scenario in which a test statistic exhibits a "discontinuity in its asymptotic distribution" and an infinite-dimensional nuisance parameter affects the asymptotic behavior of the test statistic.

In many testing problems, the asymptotic distribution of a KS-type statistic is determined by establishing the weak convergence of some underlying stochastic process and applying the continuous mapping theorem. This yields the asymptotic distribution to be the supremum of the limit process. In the context of conditional moment inequalities with drifting sequences of distributions, this method does not work. The reason is that the normalized mean function of the underlying stochastic process, i.e., $h_{1,n,F_n}(\theta_n, g)$, often (in fact, usually) does not converge uniformly over $g \in \mathcal{G}$ to its pointwise limit, i.e., $h_1(g)$, and, hence, stochastic equicontinuity fails.[48]

We show by counter-example that the asymptotic distribution under drifting sequences of null distributions of a KS statistic, where the "sup" is over $g \in \mathcal{G}$, does not necessarily equal the supremum of the limiting process indexed by $g \in \mathcal{G}$ that is determined by the finite-dimensional distributions. Hence, if the critical value is based on this limiting process, a KS test does not necessarily have correct asymptotic null rejection probability. In fact, we show that it can over-reject the null hypothesis substantially.

The same phenomenon does not arise with CvM statistics, which are "average" statistics. This is because the averaging smooths out the non-uniform convergence of the normalized mean function.

The results in the first section of this Appendix show that the problem discussed above does not arise with the KS statistic when the critical value employed is a GMS critical value that satisfies Assumption GMS1, see Section 4, or a PA critical value. The validity of these critical values is established using a uniform asymptotic approximation

[48] Note that drifting sequences of distributions are of interest because correct asymptotic coverage probabilities under all drifting sequences is necessary, though not sufficient, for correct uniform asymptotic coverage probabilities.

of the distribution of the KS statistic, rather than using asymptotics under sequences of true distributions.

To start, we give a very simple deterministic example to illustrate a situation in which a deterministic KS statistic does not converge to the supremum of the pointwise limit, but an "average" CvM statistic does converge to the average of the pointwise limit. Consider the piecewise linear functions $f_n : [0,1] \to [0,1]$ defined by

$$f_n(x) = \begin{cases} x/\varepsilon_n & \text{for } x \in [0, \varepsilon_n] \\ 1 - (x - \varepsilon_n)/\varepsilon_n & \text{for } x \in [\varepsilon_n, 2\varepsilon_n] \\ 0 & \text{for } x \in [2\varepsilon_n, 1], \end{cases} \tag{13.16}$$

where $0 < \varepsilon_n \to 0$ as $n \to \infty$. Then, for all $x \in [0,1]$,

$$f_n(x) \to f(x) = 0 \text{ as } n \to \infty. \tag{13.17}$$

The KS statistic does not converge to the supremum of the limit function:

$$\sup_{x \in [0,1]} f_n(x) = 1 \nrightarrow 0 = \sup_{x \in [0,1]} f(x) \text{ as } n \to \infty. \tag{13.18}$$

On the other hand, the CvM statistic does converge to the average of the limit function:

$$\int_0^1 f_n(x)dx = \varepsilon_n \to 0 = \int_0^1 f(x)dx \text{ as } n \to \infty. \tag{13.19}$$

The convergence result for the KS statistic in (13.18) is potentially problematic because in a testing problem with a KS statistic the critical value might be obtained from the distribution of the supremum of the limit process. If convergence in distribution of the KS statistic to the "sup" of the limit process does not hold, then such a critical value is not necessarily appropriate.

Now we show that the phenomenon illustrated in (13.16)-(13.19) arises in conditional moment inequality models. We consider a particular conditional moment inequality model with a single linear moment inequality, a fixed true value $\theta_0$, and a particular drifting sequence of distributions. (Note that CX stands for "counterexample.")

**Assumption CX.** (a) $m(W_i, \theta) = Y_i - \theta$ for $Y_i, \theta \in R$,

(b) $m(W_i, \theta_0) = Y_i = U_i + 1(X_i \in (\varepsilon_n, 1])$, where the true value $\theta_0$ equals 0, $EU_i = 0$, $EU_i^2 = 1$, the distribution of $U_i$ does not depend on $n$, $U_i$ and $X_i$ are independent, and the constants $\{\varepsilon_n : n \geq 1\}$ satisfy $\varepsilon_n \to 0$ as $n \to \infty$,

(c) $X_i = \varepsilon_n$ with probability 1/2 and $X_i$ is uniform on $[0, 1]$ with probability 1/2,

(d) $\{W_i = (Y_i, X_i)' : i \leq n, n \geq 1\}$ is a row-wise independent and identically distributed triangular array (with the dependence of $W_i$, $Y_i$, and $X_i$, on $n$ suppressed for notational simplicity),

(e) $S(m, \Sigma) = S(m)$ for $m \in R$,

(f) $S$ satisfies Assumptions S1 and S2, and

(g) $\mathcal{G} = \{g_{a,b} : g_{a,b} = 1(x \in (a, b])$ for some $0 \leq a < b \leq 1\}$.

The function $S_1(m) = [m]_-^2$ satisfies Assumptions CX(e)-(f). Assumption CX(e) is made for simplicity. It could be removed and with some changes to the proofs the results given below would hold for $S = S_2$ as well. The class of functions $\mathcal{G}$ specified in Assumption CX(g) is the class of one-dimensional boxes, as in Example 1 of Section 3.3.

We write

$$\begin{aligned} n^{1/2}\overline{m}_n(\theta_0, g_{a,b}) &= n^{-1/2}\sum_{i=1}^{n} Y_i g_{a,b}(X_i) = \nu_n(g_{a,b}) + h_{1,n}(g_{a,b}), \text{ where} \\ \nu_n(g_{a,b}) &= n^{1/2}(\overline{m}_n(\theta_0, g_{a,b}) - E_{F_n}\overline{m}_n(\theta_0, g_{a,b})) \text{ and} \\ h_{1,n}(g_{a,b}) &= n^{1/2}E_{F_n}\overline{m}_n(\theta_0, g_{a,b}). \end{aligned} \tag{13.20}$$

The KS statistic is

$$\sup_{g_{a,b}\in\mathcal{G}} S(n^{1/2}\overline{m}_n(\theta_0, g_{a,b})) = \sup_{g_{a,b}\in\mathcal{G}} S(\nu_n(g_{a,b}) + h_{1,n}(g_{a,b})).^{49} \tag{13.21}$$

Let $\nu(\cdot)$ be a mean zero Gaussian process indexed by $g_{a,b} \in \mathcal{G}$ with covariance kernel $K(\cdot, \cdot)$ and with sample paths that are uniformly $\rho$-continuous, where $K(\cdot, \cdot)$ and $\rho(\cdot, \cdot)$ are specified in the proof of Theorem B4 given in the next subsection.

The KS statistic satisfies the following result.

**Theorem B4.** *Suppose Assumption* CX *holds. Then,*

(a) $\nu_n(\cdot) \Rightarrow \nu(\cdot)$ *as* $n \to \infty$,

(b) $h_{1,n}(g_{a,b}) \to h_1(g_{a,b}) = \infty$ *as* $n \to \infty$ *for all* $g_{a,b} \in \mathcal{G}$,

(c) $\sup_{g_{a,b}\in\mathcal{G}} |h_{1,n}(g_{a,b}) - h_1(g_{a,b})| \nrightarrow 0$ *as* $n \to \infty$,

(d) $S(\nu_n(g_{a,b}) + h_{1,n}(g_{a,b})) \to_d S(\nu(g_{a,b}) + h_1(g_{a,b}))$ *as* $n \to \infty$ *for all* $g_{a,b} \in \mathcal{G}$,

(e) $\sup_{g_{a,b}\in\mathcal{G}} S(\nu(g_{a,b}) + h_1(g_{a,b})) = 0$ a.s.,

(f) $\sup_{g_{a,b}\in\mathcal{G}} S(\nu_n(g_{a,b}) + h_{1,n}(g_{a,b})) \geq S(\nu_n(g_{0,\varepsilon_n}) + h_{1,n}(g_{0,\varepsilon_n})) \to_d S(Z^*)$ *as* $n \to \infty$, *where* $Z^* \sim N(0, 1/2)$ *and the inequality holds* a.s., *and*

(g) $\sup_{g_{a,b}\in\mathcal{G}} S(\nu_n(g_{a,b}) + h_{1,n}(g_{a,b})) \not\to_d \sup_{g_{a,b}\in\mathcal{G}} S(\nu(g_{a,b}) + h_1(g_{a,b}))$ *as* $n \to \infty$.

**Comments. 1.** Theorem B4(g) shows that the KS statistic does not have an asymptotic distribution that equals the supremum over $g_{a,b} \in \mathcal{G}$ of the pointwise limit given in Theorem B4(d). This is due to the lack of uniform convergence of $h_{1,n}(g_{a,b})$ shown in Theorem B4(c). (Note that the convergence in part (d) of the Theorem also holds jointly over any finite set of $g_{a,b} \in \mathcal{G}$.)

**2.** Let $c_{\infty,1-\alpha}$ denote the $1-\alpha$ quantile of $\sup_{g_{a,b}\in\mathcal{G}} S(\nu(g_{a,b})+h_1(g_{a,b}))$. By Theorem B4(e), $c_{\infty,1-\alpha} = 0$. Theorem B4(f) and some calculations (given in the proof of Theorem B4 below) yield

$$\liminf_{n\to\infty} P\left(\sup_{g_{a,b}\in\mathcal{G}} S(\nu_n(g_{a,b}) + h_{1,n}(g_{a,b})) > c_{\infty,1-\alpha}\right) \geq 1/2. \tag{13.22}$$

That is, if one uses $c_{\infty,1-\alpha}$ as the critical value, the nominal level $\alpha$ test based on the KS statistic has an asymptotic null rejection probability that is bounded below by 1/2, which indicates substantial over-rejection.

Next, we provide results for a CvM statistic defined by

$$\int S(n^{1/2}\overline{m}_n(\theta_0, g_{a,b}))dQ(g_{a,b}) = \int S(\nu_n(g_{a,b}) + h_{1,n}(g_{a,b}))dQ(g_{a,b}), \tag{13.23}$$

where $Q$ is a probability measure on $\mathcal{G}$. In contrast to the KS statistic, the CvM statistic is well-behaved asymptotically.

**Theorem B5.** *Suppose Assumption* CX *holds. Then,*

$$\int S(\nu_n(g_{a,b}) + h_{1,n}(g_{a,b}))dQ(g_{a,b}) \to_d \int S(\nu(g_{a,b}) + h_1(g_{a,b}))dQ(g_{a,b}) \text{ as } n \to \infty.$$

**Comment.** Theorem B5 is not proved using the continuous mapping theorem due to the non-uniform convergence of $h_{1,n}(g_{a,b})$. Rather, it is proved using an almost sure representation argument coupled with the bounded convergence theorem.

## 13.5 Problems with Pointwise Asymptotics

In the case of unconditional moment inequalities, pointwise asymptotics have been shown in Andrews and Guggenberger (2009) to be deficient in the sense that they fail to capture the finite-sample properties of a typical test statistic of interest. This is due to the discontinuity in the asymptotic distribution of the test statistic. In the case of *conditional* moment equalities, the deficiency of pointwise asymptotics is even greater. We show in a simple example that the asymptotic distribution of a test statistic $T_n(\theta_0)$ under a fixed distribution $F_0$ often is *pointmass at zero* even when the true parameter $\theta_0$ is on the boundary of the identified set. This does not reflect the statistic's finite-sample distribution.

Suppose (i) $W_i = (Y_i, X_i)'$, (ii) there is one moment inequality function $m(W_i, \theta) = Y_i - \theta$ and no moment equalities (i.e., $p = 1$ and $v = 0$), (iii) the true distribution is $F_0$ for all $n \geq 1$, (iv) $Y_i = \theta_0 + \mu(X_i) + U_i$, where $X_i, U_i \in R$ and $\mu(\cdot) = \mu_{F_0}(\cdot)$, (v) $\mu(x) \geq 0$ $\forall x \in R$, $\mathcal{X}_{zero} = \{x \in Supp_{F_0}(X_i) : \mu(x) = 0\} \neq \varnothing$, and $\mu(\cdot)$ is continuous on $R$, and (vi) under $F_0$, $(X_i, U_i)$ are i.i.d., $X_i$ and $U_i$ are independent, $E_{F_0} U_i = 0$, $Var_{F_0}(U_i) = 1$, $X_i$ is absolutely continuous, and $Var_{F_0}(X_i) \in (0, \infty)$. As defined, the conditional moment inequality is

$$E_{F_0}(m(W_i, \theta_0)|X_i) = \mu(X_i) \geq 0 \text{ a.s.} \tag{13.24}$$

The inequality in (13.24) is strict except when $X_i \in \mathcal{X}_{zero}$. Often, the latter occurs with probability zero. For example, this is true if $\mathcal{X}_{zero}$ is a singleton (or a set with Lebesgue measure zero). In spite of the moment inequality being strict with probability one, the true value $\theta_0$ is on the boundary of the identified set $\Theta_{F_0}$, i.e., $\Theta_{F_0} = (-\infty, \theta_0]$.[50]

We consider a test statistic based on $S(n^{1/2}\overline{m}_n(\theta, g), I)$ with $S = S_1 = S_2$:

$$\begin{aligned} T_n(\theta_0) &= \int [n^{1/2}\overline{m}_n(\theta_0, g)]_-^2 dQ(g) \\ &= \int \left[ n^{1/2} \left( n^{-1} \sum_{i=1}^n (U_i + \mu(X_i))g(X_i) - \Delta(g) \right) + n^{1/2}\Delta(g) \right]_-^2 dQ(g), \text{ where} \end{aligned}$$

$$\overline{m}_n(\theta_0, g) = n^{-1} \sum_{i=1}^n (Y_i - \theta_0) g(X_i) \text{ and } \Delta(g) = E_{F_0}\mu(X_i)g(X_i). \tag{13.25}$$

[50] This holds because, for any $\theta > \theta_0$, (a) $E_{F_0}(m(W_i, \theta)|X_i) = \mu(X_i) + \theta_0 - \theta$, (b) $\forall \delta > 0$, $P_{F_0}(X_i \in B(\mathcal{X}_{zero}, \delta)) > 0$ by the absolute continuity of $X_i$, where $B(\mathcal{X}_{zero}, \delta)$ denotes the closed set of points that are within $\delta$ of the set $\mathcal{X}_{zero}$, (c) for $\delta^* > 0$ sufficiently small, $\mu(x) < \theta - \theta_0$ $\forall x \in B(\mathcal{X}_{zero}, \delta^*)$ by the continuity of $\mu(\cdot)$, and, hence, (d) $0 < P_{F_0}(X_i \in B(\mathcal{X}_{zero}, \delta^*)) \leq P_{F_0}(E_{F_0}(m(W_i, \theta)|X_i) < 0)$, which implies that $\theta \notin \Theta_{F_0}$.

The first summand in the integrand in (13.25) is $O_p(1)$ uniformly over $g \in \mathcal{G}$ by a functional central limit theorem (CLT) and is identically zero if $P_{F_0}(g(X_i) = 0) = 1$. The second summand, $n^{1/2}\Delta(g)$, diverges to infinity unless $\Delta(g) = 0$. In addition, $[x_n]_-^2 \to 0$ as $x_n \to \infty$. Hence, if $\Delta(g) > 0$, the integrand converges in probability to zero. In the leading case in which $\mathcal{X}_{zero}$ is a singleton set (or any set with Lebesgue measure zero), $\Delta(g) = 0$ only if $P_{F_0}(g(X_i) = 0) = 1$ (using the absolute continuity of $X_i$). In consequence, if $\Delta(g) = 0$, the integrand in (13.25) equals zero a.s. Combining these results shows that the asymptotic distribution of $T_n(\theta_0)$ under the fixed distribution $F_0$ is pointmass at zero even though the true parameter is on the boundary of the identified set.[51]

The pointmass asymptotic distribution of $T_n(\theta_0)$ does not mimic its finite-sample distribution well at all. In finite samples, the distribution of $T_n(\theta_0)$ is non-degenerate because the quantity $n^{1/2}\Delta(g)$ is finite and far from infinity for all functions $g$ for which $\mu(x)$ is not large for $x \in Supp(g)$. Pointwise asymptotics fail to capture this.

The implication of the discussion above is that to obtain asymptotic results that mimic the finite-sample situation it is necessary to consider uniform asymptotics or, at least, asymptotics under drifting sequences of distributions.

## 13.6 Subsampling Critical Values

### 13.6.1 Definition

Here we define subsampling critical values and CS's. Let $b$ denote the subsample size when the full sample size is $n$. We assume $b \to \infty$ and $b/n \to 0$ as $n \to \infty$. The number of different subsamples of size $b$ is $q_n$. There are $q_n = n!/\left(b!\,(n-b)!\right)$ different subsamples of size $b$.

Let $\{T_{n,b,j}(\theta) : j = 1, ..., q_n\}$ be subsample statistics where $T_{n,b,j}(\theta)$ is defined exactly the same as $T_n(\theta)$ is defined but based on the $j$th subsample rather than the full sample. The empirical distribution function and the $1-\alpha$ quantile of $\{T_{n,b,j}(\theta) : j = 1, ..., q_n\}$

[51] This argument is only heuristic. The result can be proved formally using a combination of an almost sure representation result and the bounded convergence theorem as in the proofs given in Supplemental Appendix A.

are

$$U_{n,b}(\theta, x) = q_n^{-1} \sum_{j=1}^{q_n} 1(T_{n,b,j}(\theta) \leq x) \text{ for } x \in R \text{ and}$$
$$c_{n,b}(\theta, 1-\alpha) = \inf\{x \in R : U_{n,b}(\theta, x) \geq 1-\alpha\}, \tag{13.26}$$

respectively. The subsampling critical value is $c_{n,b}(\theta_0, 1-\alpha)$. The nominal level $1-\alpha$ CS is given by (2.5) with $c_{n,1-\alpha}(\theta) = c_{n,b}(\theta, 1-\alpha)$.[52]

### 13.6.2 Asymptotic Coverage Probabilities of Subsampling Confidence Sets

Next, we show that nominal $1-\alpha$ subsampling CS's have asymptotic coverage probabilities greater than or equal to $1-\alpha$ under drifting sequences of parameters and distributions $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$. The sequences that we consider are those in the set $Seq^b$, which is defined as follows.

Let $\mathcal{H}_1, \mathcal{H}_2$, and $\mathcal{H}$ be defined as in (5.5). Let $\mathcal{H}_1^*(h_1) = \{h_1^* \in \mathcal{H}_1 : h_{1,j}^*(g) > 0$ only if $h_{1,j}(g) = \infty$ for $j \leq p, \forall g \in \mathcal{G}\}$.

**Definition $\mathbf{Seq}^b(\mathbf{h}_1^*, \mathbf{h})$.** For $h \in \mathcal{H}$ and $h_1^* \in \mathcal{H}_1^*(h_1)$, define $Seq^b(h_1^*, h)$ to be the set of sequences $\{(\theta_n, F_n) : n \geq 1\}$ such that

(i) $(\theta_n, F_n) \in \mathcal{F}\ \forall n \geq 1$,

(ii) $\lim_{n\to\infty} h_{1,n,F_n}(\theta_n, g) = h_1(g)\ \forall g \in \mathcal{G}$,

(iii) $\lim_{n\to\infty} \sup_{g,g^* \in \mathcal{G}} ||D_{F_n}^{-1/2}(\theta_n)\Sigma_{F_n}(\theta_n, g, g^*)D_{F_n}^{-1/2}(\theta_n) - h_2(g, g^*)|| = 0$, and

(iv) $\lim_{n\to\infty} b^{1/2} D_{F_n}^{-1/2}(\theta_n) E_{F_n} m(W, \theta_n, g) = h_1^*(g)\ \forall g \in \mathcal{G}$.

Let

$$Seq^b = \bigcup_{h_1^* \in \mathcal{H}_1^*(h),\ h \in \mathcal{H}} Seq^b(h_1^*, h). \tag{13.27}$$

We use the following assumptions.

**Assumption SQ.** For all functions $h_1 : \mathcal{G} \to R_{[+\infty]}^p \times \{0\}^v$, $h_2 : \mathcal{G}^2 \to \mathcal{W}$, and mean zero Gaussian processes $\{\nu_{h_2}(g) : g \in \mathcal{G}\}$ with finite-dimensional covariance matrix $h_2(g, g^*)$

[52] The subsampling critical value defined above is a non-recentered subsampling critical value. One also could consider recentered subsampling critical values, see Andrews and Soares (2010) for the definition. But, there is little reason to do so because tests based on recentered subsampling critical values have the same first-order asymptotic power properties as PA tests and recentered bootstrap tests and worse behavior than the latter two tests in terms of the magnitude of errors in null rejection probabilities asymptotically.

for $g, g^* \in \mathcal{G}$, the distribution function of $\int S(\nu_{h_2}(g)+h_1(g), h_2(g)+\varepsilon I_k)dQ(g)$ at $x \in R$ is

(a) continuous for $x > 0$ and

(b) strictly increasing for $x > 0$ unless $v = 0$ and $h_1(g) = \infty^p$ a.s. $[Q]$.

Lemma B3 below shows that Assumption SQ is satisfied by $S_1$ and $S_2$.

**Lemma B3.** *Assumption* SQ *holds when* $S = S_1$ *or* $S_2$.

The following Assumption C is needed only to show that subsampling CS's are not asymptotically conservative. For $(\theta, F) \in \mathcal{F}$, define $h_{1,j,F}(\theta, g) = \infty$ if $E_F m_j(W_i, \theta, g) > 0$ and $h_{1,j,F}(\theta, g) = 0$ if $E_F m_j(W_i, \theta, g) = 0$ for $g \in \mathcal{G}, j = 1, ..., p$. Let $h_{1,F}(\theta, g) = (h_{1,1,F}(\theta, g), ..., h_{1,p,F}(\theta, g), 0_v')'$.

**Assumption C.** For some $(\theta, F) \in \mathcal{F}$, $\int S(\nu_{h_{2,F}}(\theta, g)+h_{1,F}(\theta, g), h_{2,F}(\theta, g)+\varepsilon I_k)dQ(g)$ is continuous at its $1-\alpha$ quantile, where $\{\nu_{h_{2,F}}(\theta, g) : g \in \mathcal{G}\}$ is a mean zero Gaussian process concentrated on the space of uniformly $\rho$-continuous bounded $R^k$-valued functionals on $\mathcal{G}$, i.e., $U_\rho^k(\mathcal{G})$, with covariance kernel $h_{2,F}(\theta, g, g^*)$ for $g, g^* \in \mathcal{G}$.

Assumption C is not very restrictive.

The exact and asymptotic confidence sizes of a subsampling CS are

$$ExCS_n = \inf_{(\theta,F)\in\mathcal{F}} P_F\left(T_n(\theta) \le c_{n,b}(\theta, 1-\alpha)\right) \text{ and } AsyCS = \liminf_{n\to\infty} ExCS_n. \quad (13.28)$$

The next assumption is used to establish $AsyCS$ for subsampling CS's. It is a high-level condition that is difficult to verify and hence is not very satisfactory.

**Assumption Sub.** For some subsequence $\{v_n : n \ge 1\}$ of $\{n\}$ for which $\{(\theta_{v_n}, F_{v_n}) \in \mathcal{F} : n \ge 1\}$ satisfies $\lim_{n\to\infty} P_{F_{v_n}}(T_n(\theta_{v_n}) \le c_{n,b}(\theta_{v_n}, 1-\alpha)) = AsyCS$ (such a subsequence always exists), there is a subsequence $\{m_n\}$ of $\{v_n\}$ such that $\{(\theta_{m_n}, F_{m_n}) \in \mathcal{F} : n \ge 1\}$ belongs to $Seq^b$, where $Seq^b$ is defined with $m_n$ in place of $n$ throughout.

Part (a) of the following Theorem shows that subsampling CS's have correct asymptotic coverage probabilities under drifting sequences of parameters and distributions.

**Theorem B6.** *Suppose Assumptions* M, S1, S2, *and* SQ *hold. Then, a nominal* $1-\alpha$ *subsampling confidence set based on* $T_n(\theta)$ *satisfies*

(a) $\inf_{\{(\theta_n,F_n):n\ge1\}\in Seq^b} \liminf_{n\to\infty} P_{F_n}(T_n(\theta_n) \le c_{n,b}(\theta_n, 1-\alpha)) \ge 1-\alpha$,

(b) *if Assumption* C *also holds, then*

$$\inf_{\{(\theta_n,F_n):n\geq 1\}\in Seq^b} \liminf_{n\to\infty} P_{F_n}(T_n(\theta_n) \leq c_{n,b}(\theta_n, 1-\alpha)) = 1-\alpha, \textit{ and}$$

(c) *if Assumptions* Sub *and* C *also hold, then* $AsyCS = 1-\alpha$.

**Comment.** Theorem B6(c) establishes that subsampling CS's have correct $AsyCS$ provided Assumption Sub holds. The latter condition is difficult to verify. Hence, this result is not nearly as useful as the uniformity results given for GMS and PA CS's in Section 5.

# 14 Supplemental Appendix C

In this Appendix, we prove all the results stated in the main paper except for Theorems 1 and 2(a), which are proved in Supplemental Appendix A, and Lemma A1, which is proved in Supplemental Appendix E. The proofs are given in the following order: Lemma 2, Lemma 3, Theorem 2(b), Lemma 4, Theorem 3, Theorem 4, and Lemma 1.

## 14.1 Proofs of Lemmas 2 and 3 and Theorem 2(b)

**Proof of Lemma 2.** We have: $\theta \notin \Theta_F(\mathcal{G})$ implies that $E_F m_j(W_i, \theta) g_j(X_i) < 0$ for some $j \leq p$ or $E_F m_j(W_i, \theta) g_j(X_i) \neq 0$ for some $j = p+1, ..., k$. By the law of iterated expectations and $g_j(x) \geq 0$ for all $x \in R^{d_x}$ and $j \leq p$, this implies that $P_F(X_i \in \mathcal{X}_F(\theta)) > 0$ and, hence, $\theta \notin \Theta_F$.

On the other hand, $\theta \notin \Theta_F$ implies that $P_F(X_i \in \mathcal{X}_F(\theta)) > 0$ and the latter implies that $\theta \notin \Theta_F(\mathcal{G})$ by Assumption CI. $\square$

The proof of Lemma 3 uses the following Lemma, which is an existence and uniqueness result. The proof of the Lemma utilizes an extended measure result from Billingsley (1995, Thm. 11.3), which delivers the existence part of the Lemma. The proof is given after the proof of Lemma 3.

**Lemma C1.** *Let $\mathcal{R}$ be a semi-ring of subsets of $R^{d_x}$. Let $\mu$ be a bounded countably additive set function on $\sigma(\mathcal{R})$ such that $\mu(\phi) = 0$ and $\mu(C) \geq 0$ for all $C \in \mathcal{R} \cup \{R^{d_x}\}$. If $R^{d_x}$ can be written as the union of a countable number of disjoint sets in $\mathcal{R}$, then $\mu$ is a measure on $\sigma(\mathcal{R})$ (and hence $\mu(C) \geq 0$ for all $C \in \sigma(\mathcal{R})$).*[53]

**Proof of Lemma 3.** First, we establish Assumption CI for $\mathcal{G} = \mathcal{G}_{box}$ with $\bar{r} = \infty$. It suffices to show

$$\begin{aligned} E_F(m_j(W_i, \theta) g_j(X_i)) \geq 0 \ \forall g \in \mathcal{G} &\Rightarrow E_F(m_j(W_i, \theta)|X_i) \geq 0 \text{ a.s.} \\ &\qquad \text{for } j = 1, ..., p \text{ and} \\ E_F(m_j(W_i, \theta) g_j(X_i)) = 0 \ \forall g \in \mathcal{G} &\Rightarrow E_F(m_j(W_i, \theta)|X_i) = 0 \text{ a.s.} \\ &\qquad \text{for } j = p+1, ..., k. \end{aligned} \tag{14.1}$$

[53] A class of subsets, $\mathcal{R}$, of a universal set is called a semi-ring if (a) the empty set $\phi \in \mathcal{R}$; (b) $A, B \in \mathcal{R}$ implies $A \cap B \in \mathcal{R}$; (c) if $A, B \in \mathcal{R}$ and $A \subset B$, then there exist disjoint sets $C_1, ..., C_N \subset \mathcal{R}$ such that $B - A = \bigcup_{i=1}^{N} C_i$, see Billingsley (1995, p.138).

We use the following set function:

$$\mu_j(C) = \sigma_{F,j}^{-1}(\theta) E_F m_j(W_i, \theta) 1(X_i \in C) \text{ for } C \in \sigma(\mathcal{C}_{box}) = \mathcal{B}(R^{d_x}), \tag{14.2}$$

where $\sigma(\mathcal{C}_{box})$ denotes the $\sigma$-field generated by $\mathcal{C}_{box}$, $\mathcal{B}(R^{d_x})$ is the Borel $\sigma$-field on $R^{d_x}$, and $\sigma(\mathcal{C}_{box}) = \mathcal{B}(R^{d_x})$ is a well-known result. First we show $\mu_j(R^{d_x}) \geq 0$. Let $I_L = (-L, L]^{d_x}$. Then, $I_L \in \mathcal{C}_{box}$ and $\mu_j(I_L) \geq 0$. We have

$$\begin{aligned} 0 \leq \lim_{L \to \infty} \mu_j(I_L) &= \lim_{L \to \infty} \sigma_{F,j}^{-1}(\theta) E_F m_j(W_i, \theta) 1(X_i \in I_L) \\ &= \sigma_{F,j}^{-1}(\theta) E_F m_j(W_i, \theta) 1(X_i \in R^{d_x}) = \mu_j(R^{d_x}), \end{aligned} \tag{14.3}$$

where the second equality holds by the dominated convergence theorem, $\sigma_{F,j}^{-1}(\theta) m_j(w, \theta) \times 1(x \in I_L) \to \sigma_{F,j}^{-1}(\theta) m_j(w, \theta) 1(x \in R^{d_x})$ as $L \to \infty$, $|\sigma_{F,j}^{-1}(\theta) m_j(w, \theta) 1(x \in I_L)| \leq \sigma_{F,j}^{-1}(\theta) |m_j(w, \theta)|$ for all $w$, and $\sigma_{F,j}^{-1}(\theta) E_F |m_j(W_i, \theta)| < \infty$.

Next, we treat the cases $j \leq p$ and $j > p$ separately because different techniques are employed. First, we consider $j = 1, ..., p$. Suppose $E_F m_j(W_i, \theta) g_j(X_i) \geq 0 \; \forall g \in \mathcal{G}$. Then, $\mu_j(C) \geq 0 \; \forall C \in \mathcal{C}_{box}$. We want to show that $E_F m_j(W_i, \theta) 1(X_i \in C) \geq 0 \; \forall C \in \mathcal{B}(R^{d_x})$ because this implies that $E_F(m_j(W_i, \theta)|X_i) \geq 0$ a.s. since $X_i$ is Borel measurable.

By Lemma C1, we have $\mu_j(C) \geq 0 \; \forall C \in \sigma(\mathcal{C}_{box})$ if (a) $\mathcal{C}_{box}$ is a semi-ring of subsets of $R^{d_x}$, (b) $\mu_j$ is bounded, (c) $\mu_j$ is countably additive, (d) $\mu_j(\phi) = 0$, (e) $\mu_j(R^{d_x}) \geq 0$, and (f) $R^{d_x}$ can be written as the union of a countable number of disjoint sets in $\mathcal{C}_{box}$. It is a well-known result that (a) holds (provided $\phi$ is added to $\mathcal{C}_{box}$). By condition (vi) in (2.3), (b) holds. Condition (c) holds by the dominated convergence theorem. Because $1(X_i \in \phi) = 0$, condition (d) holds. By (14.3), condition (e) holds. Condition (f) holds because

$$R^{d_x} = \bigcup_{\{i_1, i_2, ..., i_k\} \in \mathbb{N}^k} \prod_{j=1}^{k} (i_j, i_j + 1], \tag{14.4}$$

where $\mathbb{N}$ is the set of all natural numbers. Therefore, $\mu_j(C) \geq 0 \; \forall C \in \sigma(\mathcal{C}_{box}) = \mathcal{B}(R^{d_x})$, i.e.,

$$E_F m_j(W_i, \theta) 1(X_i \in C) \geq 0 \; \forall C \in \mathcal{B}(R^{d_x}). \tag{14.5}$$

Next, we consider $j = p + 1, ..., k$. Suppose $E_F m_j(W_i, \theta) g_j(X_i) = 0 \; \forall g \in \mathcal{G}_{box}$. Then, $\mu_j(C) = 0 \; \forall C \in \mathcal{C}_{box}$ We want to show that $E_F m_j(W_i, \theta) 1(X_i \in C) = 0 \; \forall C \in \mathcal{B}(R^{d_x})$ because this implies that $E_F(m_j(W_i, \theta)|X_i) = 0$ a.s. because $X_i$ is Borel measurable. To do so, we show that $\mathcal{C}_0 = \mathcal{B}(R^{d_x})$, where $\mathcal{C}_0 \equiv \{C \in \mathcal{B}(R^{d_x}) : \mu_j(C) = 0\}$. It suffices to show $\mathcal{B}(R^{d_x}) \subset \mathcal{C}_0$. Because $\mathcal{C}_{box} \subset \mathcal{C}_0$ and $\sigma(\mathcal{C}_{box}) = \mathcal{B}(R^{d_x})$, it suffices to show that $\mathcal{C}_0$ is a $\sigma$-field. The set $\mathcal{C}_0$ is indeed a $\sigma$-field because (a) $R^{d_x} \in \mathcal{C}_0$ by (14.3), (b) if $C \in \mathcal{C}_0$, then $\mu_j(C^c) = \mu_j(R^{d_x}) - \mu_j(C) = 0$, i.e., $C^c \in \mathcal{C}_0$, and (c) if $C_1, C_2, ...$ are disjoint sets in $\mathcal{C}_0$, then $\mu_j(\bigcup_{i=1}^{\infty} C_i) = \sum_{i=1}^{\infty} \mu_j(C_i) = 0$ because $\mu_j$ is an additive set function, i.e., $\bigcup_{i=1}^{\infty} C_i \in \mathcal{C}_0$. This completes the proof of Assumption CI for $\mathcal{G} = \mathcal{G}_{box}$ with $\bar{r} = \infty$.

Assumption CI holds for $\mathcal{G} = \mathcal{G}_{box}$ with $\bar{r} = \infty$ implies that Assumption CI holds for $\mathcal{G} = \mathcal{G}_{box}$ when $\bar{r} \in (0, \infty)$. The reason is that if some deviation is captured by a big box, it also must be captured by some smaller box contained in the big box (because a big box is a finite disjoint union of smaller boxes).

For $\mathcal{G} = \mathcal{G}_{c\text{-}cube}$, Assumption CI holds by the same argument as for $\mathcal{G}_{box}$ but with $\mathcal{C}_{c\text{-}cube}$ in place of $\mathcal{C}_{box}$ provided (i) $\mathcal{C}_{c\text{-}cube} \cup \{\phi\}$ is a semi-ring of subsets of $[0,1]^{d_x}$, (ii) $[0,1]^{d_x}$ can be written as the union of a countable number of disjoint sets in $\mathcal{C}_{c\text{-}cube}$, and (iii) $\sigma(\mathcal{C}_{c\text{-}cube}) = \mathcal{B}([0,1]^{d_x})$. Condition (i) is straightforward to verify. Condition (ii) is verified by using $\cup_{\ell=1}^{2r}((\ell-1)/(2r), \ell/(2r)] = [0,1]$ (since the interval $(0, 1/(2r)]$ is defined specially to include 0) to construct a finite number of $d_x$-dimensional boxes whose union is $[0,1]^{d_x}$. Condition (iii) holds because every element of $\mathcal{C}_{box}$ can be written as a countable union of sets in $\mathcal{C}_{c\text{-}cube}$ and $\sigma(\mathcal{C}_{box}) = \mathcal{B}([0,1]^{d_x})$.

Finally, we establish Assumption M. For $\mathcal{G} = \mathcal{G}_{box}$, Assumptions M(a) and M(b) hold by taking $G(x) = 1\ \forall x$ and $\delta_1 = 4/\delta + 3$. Assumption M(c) holds because $\mathcal{C}_{box}$ forms a Vapnik-Cervonenkis class of sets. Assumption M holds for $\mathcal{G}_{c\text{-}cube}$ because $\mathcal{G}_{c\text{-}cube} \subset \mathcal{G}_{box}$. $\square$

**Proof of Lemma C1.** Because (i) $\mu : \sigma(\mathcal{R}) \to R$ is a bounded countably additive set function, (ii) $\mu(\phi) = 0$, and (iii) $\mu(C) \geq 0\ \forall C \in \mathcal{R}$, Billingsley's (1995) Thm. 11.3 implies that there exist a measure, $\mu^*$, on $\sigma(\mathcal{R})$ that agrees with $\mu$ on $\mathcal{R}$. We want to show that $\mu^*$ agrees with $\mu$ on $\sigma(\mathcal{R})$. That is, we want to show that $\mathcal{C}_{eq} = \sigma(\mathcal{R})$, where

$$\mathcal{C}_{eq} = \{C \in \sigma(\mathcal{R}) : \mu^*(C) = \mu(C)\}. \tag{14.6}$$

It suffices to show that $\sigma(\mathcal{R}) \subseteq \mathcal{C}_{eq}$ because by definition, $\sigma(\mathcal{R}) \supseteq \mathcal{C}_{eq}$. We use Dynkin's $\pi$-$\lambda$ theorem, e.g., see Billingsley (1995, p.33), to establish this.

Because $\mathcal{R}$ is a semi-ring, $\mathcal{R}$ is a $\pi$-system. Now, we show that $\mathcal{C}_{eq}$ is a $\lambda$-system. By definition, the set $\mathcal{C}_{eq}$ is a $\lambda$-system if (a) $R^{d_x} \in \mathcal{C}_{eq}$, (b) $\forall C_1, C_2 \in \mathcal{C}_{eq}$ such that $C_1 \subset C_2$, $C_2 - C_1 \in \mathcal{C}_{eq}$, and (c) $\forall C_1, C_2, ... \in \mathcal{C}_{eq}$ such that $C_i \uparrow C$, $C \in \mathcal{C}_{eq}$. We show (a), (b), and (c) in turn.

(a) By assumption, $R^{d_x}$ can be written as the union of countable disjoint $\mathcal{R}$-sets, say $C_1, C_2, ... \in \mathcal{R}$, where $R^{d_x} = \bigcup_{i=1}^{n} C_i$. By countable additivity of both $\mu$ and $\mu^*$, we have

$$\mu(R^{d_x}) = \sum_{i=1}^{\infty} \mu(C_i) = \sum_{i=1}^{\infty} \mu^*(C_i) = \mu^*(R^{d_x}), \tag{14.7}$$

where the second equality holds because $C_1, C_2, ... \in \mathcal{R}$ and $\mu^*$ agrees with $\mu$ on $\mathcal{R}$. Thus condition (a) holds.

(b) Suppose $C_1, C_2 \in \mathcal{C}_{eq}$ and $C_1 \subset C_2$, then $C_2 = (C_2 - C_1) \cup C_1$. Thus,

$$\mu(C_2 - C_1) = \mu(C_2) - \mu(C_1) = \mu^*(C_2) - \mu^*(C_1) = \mu^*(C_2 - C_1), \tag{14.8}$$

where the first and the third equalities hold by the countable additivity of $\mu$ and $\mu^*$ and the second equality holds because $C_1, C_2 \in \mathcal{C}_{eq}$. Thus, condition (b) holds.

(c) Suppose $C_1, C_2, ... \in \mathcal{C}_{eq}$ and $C_i \uparrow C$, then $C = C_1 \cup (\bigcup_{i=2}^{\infty}(C_i - C_{i-1}))$ and $C_1, C_2 - C_1, ...$ are mutually disjoint. By condition (b), $C_i - C_{i-1} \in \mathcal{C}_{eq}$ for $i \geq 2$. Thus,

$$\mu(C) = \mu(C_1) + \sum_{i=2}^{\infty} \mu(C_i - C_{i-1}) = \mu^*(C_1) + \sum_{i=2}^{\infty} \mu^*(C_i - C_{i-1}) = \mu^*(C). \tag{14.9}$$

That is, condition (c) holds.

Therefore, $\mathcal{C}_{eq}$ is a $\lambda$-system. Because $\mathcal{R} \subset \mathcal{C}_{eq}$ by Dynkin's $\pi$-$\lambda$ theorem, $\sigma(\mathcal{R}) \subseteq \mathcal{C}_{eq}$. In consequence, $\sigma(\mathcal{R}) = \mathcal{C}_{eq}$, i.e., $\mu^*$ agrees with $\mu$ on $\sigma(\mathcal{R})$. Because $\mu^*$ is a measure on $\sigma(\mathcal{R})$, $\mu$ must be a measure on $\sigma(\mathcal{R})$. $\square$

**Proof of Theorem 2(b).** Consider the parameters $(\theta_c, F_c)$ that appear in Assumption GMS2. First, we determine the asymptotic behavior of the critical value $c(\varphi_n(\theta_c), \widehat{h}_{n,2}(\theta_c), 1-\alpha)$ under $(\theta_c, F_c)$. We have

$$\begin{aligned} \xi_n(\theta_c, g) &= \kappa_n^{-1} n^{1/2} \overline{D}_n^{-1/2}(\theta_c, g) \overline{m}_n(\theta_c, g) \\ &= \overline{D}_n^{-1/2}(\theta_c, g) D_{F_c}^{1/2}(\theta_c) \kappa_n^{-1} [\nu_{n,F_c}(\theta_c, g) + h_{1,n,F_c}(\theta_c, g)] \\ &= Diag^{-1/2}(\overline{h}_{2,n,F_c}(\theta_c, g)) \kappa_n^{-1} [\nu_{n,F_c}(\theta_c, g) + h_{1,n,F_c}(\theta_c, g)]. \end{aligned} \tag{14.10}$$

Note that $\overline{h}_{2,n,F_c}(\theta_c, g)$ is a function of $\widehat{h}_{2,n,F_c}(\theta_c, g, g)$ by (5.2). Let

$$T_n^{GMS}(\theta_c) = \int S(\nu_{\widehat{h}_{2,n}(\theta_c)}(g) + \varphi_n(\theta_c, g), \widehat{h}_{2,n}(\theta_c, g) + \varepsilon I_k) dQ(g). \tag{14.11}$$

Equations (4.10), (12.26), (14.10), and (14.11) imply that the distribution of $T_n^{GMS}(\theta_c)$ is determined by the joint distribution of $\{\nu_{\widehat{h}_{2,n}(\theta_c)}(g) : g \in \mathcal{G}\}$, $\{\widehat{h}_{2,n,F_c}(\theta_c, g) : g \in \mathcal{G}\}$, and $\{\kappa_n^{-1}\nu_{n,F_c}(\theta_c, g) : g \in \mathcal{G}\}$.

We have $\{(\theta_c, F_c) : n \geq 1\} \in SubSeq(h_{2,F_c}(\theta_c))$ because $(\theta_c, F_c) \in \mathcal{F}$. Hence, by Lemma A1(b), $d(\widehat{h}_{2,n,F_c}(\theta_c), h_{2,F_c}(\theta_c)) \to_p 0$ as $n \to \infty$. By the same argument as in (12.26), this yields $d(\widehat{h}_{2,n}(\theta_c), h_{2,F_c}(\theta_c)) \to_p 0$. The latter, the independence of $\widehat{h}_{2,n,F_c}(\theta_c)$ and $\{\nu_{h_2}(\cdot) : h_2 \in \mathcal{H}_2\}$, and an almost sure representation argument imply that the Gaussian processes $\{\nu_{\widehat{h}_{2,n}(\theta_c)}(\cdot) : n \geq 1\}$ converge weakly to $\nu_{h_{2,F_c}(\theta_c)}(\cdot)$ as $n \to \infty$. The sequence of random processes $\{\widehat{h}_{2,n}(\theta_c, \cdot) : n \geq 1\}$ converges in probability uniformly (and hence in distribution) to $h_{2,F_c}(\theta_c, \cdot)$, where $\widehat{h}_{2,n}(\theta_c, g) = \widehat{h}_{2,n}(\theta_c, g, g)$ and $h_{2,F_c}(\theta_c, g) = h_{2,F_c}(\theta_c, g, g)$. The sequence $\{\kappa_n^{-1}\nu_{n,F_c}(\theta_c, \cdot) : n \geq 1\}$ converges in probability to zero uniformly over $g \in \mathcal{G}$ because $\kappa_n \to \infty$ and $\{\nu_{n,F_n}(\theta_c, \cdot) : n \geq 1\}$ converges to a Gaussian process with sample paths that are bounded a.s. Therefore, we have

$$\begin{pmatrix} \nu_{\widehat{h}_{2,n}(\theta_c)}(\cdot) \\ \widehat{h}_{2,n}(\theta_c, \cdot) \\ \kappa_n^{-1}\nu_{n,F_c}(\theta_c, \cdot) \end{pmatrix} \Rightarrow \begin{pmatrix} \nu_{h_{2,F_c}(\theta_c)}(\cdot) \\ h_{2,F_c}(\theta_c, \cdot) \\ 0_{\mathcal{G}} \end{pmatrix} \text{ as } n \to \infty, \tag{14.12}$$

where $\widehat{h}_{2,n}(\theta_c)$ that appears in $\nu_{\widehat{h}_{2,n}(\theta_c)}(\cdot)$ is a function on $\mathcal{G} \times \mathcal{G}$ whereas $\widehat{h}_{2,n}(\theta_c, \cdot)$ is a function on $\mathcal{G}$, likewise for $\nu_{h_{2,F_c}(\theta_c)}(\cdot)$ and $h_{2,F_c}(\theta_c, \cdot)$, and $0_{\mathcal{G}}$ denotes the $R^k$-valued function on $\mathcal{G}$ that is identically $(0, ..., 0)' \in R^k$.

By the almost sure representation theorem, see Pollard (1990, Thm. 9.4), there exist $\{(\tilde{\nu}_n(g), \tilde{h}_{2,n}(g),\ \tilde{\nu}_{\kappa,n}(g)) : g \in \mathcal{G}, n \geq 1\}$ and $\{\tilde{\nu}(g), \tilde{h}_2(g) : g \in \mathcal{G}\}$ such that (i) $\{(\tilde{\nu}_n(g), \tilde{h}_{2,n}(g), \tilde{\nu}_{\kappa,n}(g)) : g \in \mathcal{G}\}$ has the same distribution as $\{(\nu_{\widehat{h}_{2,n}(\theta_c)}(g),\ \widehat{h}_{2,n}(\theta_c, g),$ $\kappa_n^{-1}\nu_{n,F_c}(\theta_c, g)) : g \in \mathcal{G}\}$ for all $n \geq 1$, (ii) $\{(\tilde{\nu}(g), \tilde{h}_2(g)) : g \in \mathcal{G}\}$ has the same distribution as $\{(\nu_{h_{2,F_c}(\theta_c)}(g), h_{2,F_c}(\theta_c, g) : g \in \mathcal{G}\}$, and

$$\text{(iii)}\ \sup_{g \in \mathcal{G}} \left\| \begin{pmatrix} \tilde{\nu}_n(g) \\ \tilde{h}_{2,n}(g) \\ \tilde{\nu}_{\kappa,n}(g) \end{pmatrix} - \begin{pmatrix} \tilde{\nu}(g) \\ \tilde{h}_2(g) \\ 0 \end{pmatrix} \right\| \to 0 \text{ a.s.} \tag{14.13}$$

Let

$$\tilde{T}_n^{GMS} = \int S(\tilde{\nu}_n(g) + \widetilde{\varphi}_n(g), \tilde{h}_{2,n}(g) + \varepsilon I_k) dQ(g), \tag{14.14}$$

where $\widetilde{\varphi}_n(g)$ is defined just as $\varphi_n(\theta, g)$ is defined in (4.10) but with $\tilde{h}_{2,n,j}(g) + \varepsilon\tilde{h}_{2,n,j}(1_k)$

in place of $\overline{h}_{2,n,F_n,j}(\theta, g)$, where $\tilde{h}_{2,n,j}(g)$ denotes the $(j, j)$ element of $\tilde{h}_{2,n}(g)$, and $\tilde{\xi}_n(g)$ in place of $\xi_n(\theta, g)$, where

$$\tilde{\xi}_n(g) = Diag(\tilde{h}_{2,n}(g) + \varepsilon\tilde{h}_{2,n}(1_k))^{-1/2}(\kappa_n^{-1}\tilde{\nu}_{\kappa,n}(g) + \kappa_n^{-1}h_{1,n,F_c}(\theta_c, g)). \tag{14.15}$$

Then, $\tilde{T}_n^{GMS}$ and $T_n^{GMS}(\theta_c)$ have the same distribution for all $n \geq 1$ and the same asymptotic distribution as $n \to \infty$. Let $\tilde{c}_n(1-\alpha)$ denote the $1-\alpha+\eta$ quantile of $\tilde{T}_n^{GMS}$ plus $\eta$, where $\eta$ is as in the definition of $c(h, 1-\alpha)$. Then, $\tilde{c}_n(1-\alpha)$ has the same distribution as $c(\varphi_n(\theta_c), \widehat{h}_{2,n}(\theta_c), 1-\alpha)$ for all $n \geq 1$.

Let $\tilde{\Omega}^*$ be the collection of $\omega \in \Omega$ such that at $\omega$, $\tilde{\nu}(g)(\omega)$ is bounded and the convergence in (14.13) holds. By (14.13) and the fact that the sample paths of $\{\tilde{\nu}(g) : g \in \mathcal{G}\}$ are bounded a.s., we have $P_{F_c}(\tilde{\Omega}^*) = 1$.

Under $(\theta_c, F_c)$ for all $n \geq 1$,

$$\kappa_n^{-1}h_{1,n,F_c}(\theta_c, g) = \kappa_n^{-1}n^{1/2}D_{F_c}^{-1/2}(\theta_c)E_{F_c}m(W_i, \theta_c, g) \to h_{1,\infty,F_c}(\theta_c, g) \tag{14.16}$$

as $n \to \infty$ using Assumption GMS2(c). Thus, for fixed $\omega \in \tilde{\Omega}^*$,

$$\begin{aligned}\tilde{\xi}_n(g)(\omega) &= Diag^{-1/2}(\tilde{h}_2(g) + \varepsilon\tilde{h}_2(1_k) + o(1))(o(1) + \kappa_n^{-1}h_{1,n,F_c}(\theta_c, g)) \\ &\to h_{1,\infty,F_c}(\theta_c, g),\end{aligned} \tag{14.17}$$

as $n \to \infty$ for all $g \in \mathcal{G}$, where $\tilde{h}_{2,j}(g)$ denotes the $(j, j)$ element of $\tilde{h}_2(g)$, using (14.13), $\tilde{h}_2(1_k) = I_k$ (which holds by (5.1) and Definition SubSeq($h_2$)), $\tilde{h}_{2,j}(g) \geq 0$, $\varepsilon > 0$.

By (14.17), Assumption GMS1(a), $B_n \to \infty$ as $n \to \infty$ (by Assumption GMS2(b)) and the fact that $h_{1,\infty,F_c}(\theta_c, g)$ equals either 0 or $\infty$ by definition, we have

$$\widetilde{\varphi}_n(g)(\omega) \to h_{1,\infty,F_c}(\theta_c, g) \text{ as } n \to \infty \tag{14.18}$$

for all $\omega \in \tilde{\Omega}^*$.

By (14.13), (14.15), (14.18), and Assumption S1(d), we have

$$\begin{aligned}&S(\tilde{\nu}_n(g) + \widetilde{\varphi}_n(g), \tilde{h}_{2,n}^*(g) + \varepsilon I_k)(\omega) \\ &\to S(\tilde{\nu}(g) + h_{1,\infty,F_c}(\theta_c, g), h_{2,F_c}(\theta_c, g) + \varepsilon I_k)(\omega)\end{aligned} \tag{14.19}$$

as $n \to \infty$ $\forall \omega \in \tilde{\Omega}^*, \forall g \in \mathcal{G}$. Now, by the argument given from (12.14) to the end of the proof of Theorem 1, the quantity on the left-hand side of (14.19) is bounded by a finite constant. This, (14.19), and the bounded convergence theorem give

$$\tilde{T}_n^{GMS} \to \tilde{T}^{GMS} = \int S(\tilde{\nu}(g) + h_{1,\infty,F_c}(\theta_c, g), h_{2,F_c}(\theta_c, g) + \varepsilon I_k) dQ(g) \tag{14.20}$$

as $n \to \infty$ a.s.

By (14.20),

$$P(\tilde{T}_n^{GMS} \leq x) \to P(\tilde{T}^{GMS} \leq x) \text{ as } n \to \infty \tag{14.21}$$

for all continuity points $x$ of the distribution of $\tilde{T}^{GMS}$. Let $\tilde{c}_0(1-\alpha)$ denote the $1-\alpha$ quantile of $\tilde{T}^{GMS}$. Let $\tilde{c}(1-\alpha) = \tilde{c}_0(1-\alpha+\eta) + \eta$, where $\eta$ is as in the definition of $c(h, 1-\alpha)$. By Assumption GMS2(a), the distribution function of $\tilde{T}^{GMS}$, which equals that of $T(h_{\infty,F_c}(\theta_c))$, is continuous and strictly increasing at $x = \tilde{c}(1-\alpha)$. Using Lemma 5 of Andrews and Guggenberger (2010), this gives

$$\begin{aligned} \tilde{c}_n(1-\alpha) &\to_p \tilde{c}(1-\alpha) \text{ and} \\ c(\varphi_n(\theta_c), \widehat{h}_{2,n}(\theta_c), 1-\alpha) &\to_p \tilde{c}(1-\alpha), \end{aligned} \tag{14.22}$$

where the second convergence result holds because $\tilde{c}_n(1-\alpha)$ and $c(\varphi_n(\theta_c), \widehat{h}_{2,n}(\theta_c), 1-\alpha)$ have the same distribution.

Next, by the same argument as used above to show (14.20), but with $\nu_{\widehat{h}_{2,n}(\theta_c)}(g)$ and $\varphi_n(\theta_c, g)$ replaced by $\nu_{n,F_c}(\theta_c, g)$ and $h_{1,n,F_c}(\theta_c, g)$, respectively, we have

$$T_n(\theta_c) \to_d T(h_{\infty,F_c}(\theta_c)) = \int S(\nu_{h_{2,F_c}(\theta_c)}(g) + h_{1,\infty,F_c}(\theta_c, g), h_{2,F_c}(\theta_c, g) + \varepsilon I_k) dQ(g), \tag{14.23}$$

where $h_{\infty,F_c}(\theta_c) = (h_{1,\infty,F_c}(\theta_c), h_{2,F_c}(\theta_c))$, $h_{1,n,F_c}(\theta_c) \to h_{1,\infty,F_c}(\theta_c)$ by straightforward calculations, and $\nu_{n,F_c}(\theta_c, \cdot) \Rightarrow \nu_{h_{2,F_c}(\theta_c)}(\cdot)$ by Lemma A1(a). Note that $T(h_{\infty,F_c}(\theta_c))$ and $\tilde{T}^{GMS}$ have the same distribution because $\nu_{h_{2,F_c}(\theta_c)}(\cdot)$ and $\tilde{\nu}(\cdot)$ have the same distribution. Thus, $\tilde{c}(1-\alpha)$ $(= \tilde{c}_0(1-\alpha+\eta)+\eta)$ is the $1-\alpha+\eta$ quantile of $T(h_{\infty,F_c}(\theta_c))$ plus $\eta > 0$.

By (14.22), (14.23), Assumption GMS2(a), and Lemma 5 of Andrews and Guggenberger (2010), for $\eta > 0$, we have

$$\begin{aligned} &\lim_{n\to\infty} P_{F_c}(T_n(\theta_c) \leq c(\varphi_n(\theta_c), \widehat{h}_{2,n}(\theta_c), 1-\alpha)) \\ &= P(T(h_{\infty,F_c}(\theta_c)) \leq \tilde{c}_0(1-\alpha+\eta)+\eta). \end{aligned} \tag{14.24}$$

The limit as $\eta \to 0$ of the right-hand side equals $1-\alpha$ because distribution functions are right-continuous and the distribution function of $T(h_{\infty,F_c}(\theta_c))$ at its $1-\alpha$ quantile equals $1-\alpha$ by Assumption GMS2(a).

Combining (14.24) and the result of Theorem 2(a), which holds for all $\eta > 0$ and hence holds when the limit as $\eta \to 0$ is taken, gives Theorem 2(b). $\square$

## 14.2 Proofs of Results for Fixed Alternatives

**Proof of Lemma 4.** First, we prove part (a). It holds immediately that $Supp(Q_a) = \mathcal{G}_{c\text{-}cube}$ because $\mathcal{G}_{c\text{-}cube}$ is countable and $Q_a$ has a probability mass function that is positive at each element in $\mathcal{G}_{c\text{-}cube}$.

Next, for part (b), consider $g = g_{x,r} \in \mathcal{G}_{box}$, where $g_{x,r}(y) = 1(y \in C_{x,r}) \cdot 1_k$ and $(x,r) \in [0,1]^{d_x} \times (0,\bar{r})^{d_x}$. Let $\delta > 0$ be given. The idea of the proof is to find a set $G_{g,\bar{\eta}} \subset \mathcal{B}_{\rho_X}(g,\delta)$ ($\subset \mathcal{G}_{box}$) such that $Q_b(G_{g,\bar{\eta}}) > 0$. This implies that $Q_b(\mathcal{B}_{\rho_X}(g,\delta)) > 0$, which is the desired result.

The set $G_{g,\bar{\eta}}$ needs to be defined differently (for reasons stated below) depending on whether $x_u < 1$ or $x_u = 1$, for $u = 1, ..., d_x$, where $x = (x_1, ..., x_{d_x})'$. For $\bar{\eta} > 0$, define

$$\begin{aligned}
G_{g,\bar{\eta}} &= \{g_{x+\eta_0,r+\eta_1} : (\eta_0,\eta_1) \in \Xi_{g,\bar{\eta}}\}, \text{ where} \\
\Xi_{g,\bar{\eta}} &= \{(\eta_0,\eta_1) \in R^{2d_x} : \text{ for } u = 1,...,d_x, \text{ if } x_u < 1, \ \eta_{0,u} \in [\bar{\eta}, 2\bar{\eta}] \ \& \\
&\qquad \eta_{1,u} \in [0,\bar{\eta}] \text{ and for } x_u = 1, \ \eta_{0,u} \in [-\bar{\eta}, 0] \ \& \ \eta_{1,u} \in [-2\bar{\eta}, -\bar{\eta}]\}. \quad (14.25)
\end{aligned}$$

We have $Q_b(G_{g,\bar{\eta}}) = Q_b^*((x,r) + \Xi_{g,\bar{\eta}}) > 0$ for all $\bar{\eta} > 0$ because $Q_b^*$ is the uniform distribution on $[0,1]^{d_x} \times (0,\bar{r})^{d_x}$.

Next, we show that $G_{g,\bar{\eta}} \subset \mathcal{B}_{\rho_X}(g,\delta)$. Let $U_{(x_u<1)} \subset \{1,...,d_x\}$ be the set of indices $u$ such that $x_u < 1$ and let $U_{(x_u=1)} \subset \{1,...,d_x\}$ be the set of indices $u$ such that $x_u = 1$. Let $g_{x+\eta_0,r+\eta_1} \in G_{g,\bar{\eta}}$. The $u$th lower endpoints of the $C_{x,r}$ and $C_{x+\eta_0,r+\eta_1}$ boxes are $x_u - r_u$ and $x_u + \eta_{0,u} - (r_u + \eta_{1,u})$, respectively. The lower endpoint of the $C_{x+\eta_0,r+\eta_1}$ box is larger than that of $C_{x,r}$ box because $\eta_{0,u} - \eta_{1,u} \in [0, 2\bar{\eta}]$ (whether $u \in U_{(x_u<1)}$ or $u \in U_{(x_u=1)}$). The $u$th upper endpoints of the $C_{x,r}$ and $C_{x+\eta_0,r+\eta_1}$ boxes are $x_u + r_u$ and $x_u + \eta_{0,u} + r_u + \eta_{1,u}$, respectively. If $u \in U_{x_u<1}$, the upper endpoint of the $C_{x+\eta_0,r+\eta_1}$ box is larger than that of $C_{x,r}$ box because $\eta_{0,u} + \eta_{1,u} \in [0, 3\bar{\eta}]$. If $u \in U_{(x_u=1)}$, the $u$th upper endpoint of the $C_{x+\eta_0,r+\eta_1}$ box is smaller than that of $C_{x,r}$ box because $\eta_{0,u} + \eta_{1,u} \in [-3\bar{\eta}, 0]$.

Using the results of the previous paragraph, we have

$$
\begin{aligned}
&\rho_X^2(g_{x,r}, g_{x+\eta_0,r+\eta_1}) \\
&= E_{F_{X,0}}[1(X_i \in C_{x,r}) - 1(X_i \in C_{x+\eta_0,r+\eta_1})]^2 \\
&\leq \sum_{u=1}^{d_x} P_{F_{X,0}}(X_{i,u} \in (x_u - r_u, x_u + \eta_{0,u} - (r_u + \eta_{1,u})]) \\
&\quad + \sum_{u \in U_{(x_u<1)}} P_{F_{X,0}}(X_{i,u} \in (x_u + r_u, x_u + \eta_{0,u} + r_u + \eta_{1,u}]) \\
&\quad + \sum_{u \in U_{(x_u=1)}} P_{F_{X,0}}(X_{i,u} \in (1 + \eta_{0,u} + r_u + \eta_{1,u}, 1 + r_u] \cap [0,1]) \\
&\leq \sum_{u=1}^{d_x} P_{F_{X,0}}(X_{i,u} \in (x_u - r_u, x_u - r_u + 2\bar{\eta}]) + \sum_{u \in U_{(x_u<1)}} P_{F_{X,0}}(X_{i,u} \in (x_u + r_u, x_u + r_u + 3\bar{\eta}]) \\
&\quad + \sum_{u \in U_{(x_u=1)}} P_{F_{X,0}}(X_{i,u} \in (1 + r_u - 3\bar{\eta}, 1 + r_u] \cap [0,1]), \qquad (14.26)
\end{aligned}
$$

where the first inequality uses the $d_x$-dimensional extension of the one-dimensional result that $(a,b]\Delta(c,d] \subset (a,c] \cup (b,d]$ when $a \leq c$ and $b \leq d$, where $\Delta$ denotes the symmetric difference of two sets.

The first and second summands on the rhs of (14.26) tend to zero as $\bar{\eta} \downarrow 0$ by the right continuity of distribution functions. The third summand on the rhs equals zero when $\bar{\eta}$ is sufficiently small (i.e., when $3\bar{\eta} < \min_{u \leq d_x} r_u$). Therefore, for $\bar{\eta} > 0$ sufficiently small, $\rho_X^2(g_{x,r}, g_{x+\eta_0,r+\eta_1}) < \delta$ and $G_{g,\bar{\eta}} \subset \mathcal{B}_{\rho_X}(g,\delta)$. This completes the proof of part (b).

Note that in the proof of part (b) we cannot treat the case where $u \in U_{(x_u=1)}$ in the same way that we treat the case for $u \in U_{(x_u<1)}$ because for $u \in U_{(x_u<1)}$ we use the center point $x_u + \eta_{0,u} > x_u$ which is not in $[0,1]$ if $x_u = 1$ and hence violates the assumption of part (b) that the centers of the $\mathcal{G}_{box}$ boxes lie in $[0,1]^{d_x}$. Conversely, we cannot treat the case where $u \in U_{(x_u<1)}$ in the same way that we treat the case for $u \in U_{(x_u=1)}$ because doing so would lead to a term $P_{F_{X,0}}(X_{i,u} \in (1 + r_u - 3\bar{\eta}, 1 + r_u])$ in (14.26) that does not go to zero as $\bar{\eta} \downarrow 0$ if $X_{i,u}$ has positive probability of equaling $1 + r_u$. $\square$

**Proof of Theorem 3.** Part (a) follows from part (b) because

$$c(\varphi_n(\theta_*), \widehat{h}_{2,n}(\theta_*), 1-\alpha) \leq c(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_*), 1-\alpha), \qquad (14.27)$$

which holds because $\varphi_n(\theta_*, g) \geq 0_k \; \forall g \in \mathcal{G}$ by Assumption GMS1(a), $c(h_1, \widehat{h}_{2,n}(\theta_*), 1-\alpha)$ is non-increasing in the first $p$ elements of $h_1$ by Assumption S1(b), and the last $v$ elements of $\varphi_n(\theta_*, g)$ equal zero.

Now we prove part (b). By Assumptions FA(a) and CI, $\beta(g_0) > 0$ for some $g_0 \in \mathcal{G}$. By construction, $e_j = m_j^*(g_0)/\beta(g_0) \in [-1, \infty)$ for $j = 1, ..., k$ and $e_j = -1$ for some $j \leq p$ or $|e_j| = 1$ for some $j = p+1, ..., k$. As defined above, $\mathcal{B}_{\rho_X}(g_0, \tau_2)$ denotes a $\rho_X$-ball centered at $g_0$ with radius $\tau_2 > 0$, where $\rho_X$ is defined in (6.3). First we show that for some $\tau_2 > 0$,

$$\int_{\mathcal{B}_{\rho_X}(g_0,\tau_2)} S\left(m^*(g)/\beta(g_0), h_{2,0}(g) + \varepsilon I_k\right) dQ(g) > 0, \qquad (14.28)$$

where $m^*(g) = (m_1^*(g), ..., m_k^*(g))'$ and $h_{2,0}(g) = h_{2,F_0}(\theta_*, g)$. We have: for $j = 1, ..., k$,

$$\begin{aligned} & |m_j^*(g) - m_j^*(g_0)| \\ &= |E_{F_0} m_j(W_i, \theta_*) g_j(X_i) - E_{F_0} m_j(W_i, \theta_*) g_{0,j}(X_i)| / \sigma_{F_0,j}(\theta_*) \\ &\leq (E_{F_0} m_j^2(W_i, \theta_*))^{1/2} (E_{F_0}(g_j(X_i) - g_{0,j}(X_i))^2)^{1/2} / \sigma_{F_0,j}(\theta_*) \\ &\leq (E_{F_0}||m(W_i, \theta_*)||^2)^{1/2} \rho_X(g, g_0) / \sigma_{F_0,j}(\theta_*), \end{aligned} \qquad (14.29)$$

where $g_{0,j}(X_i)$ denotes the $j$th element of $g_0(X_i)$.

Given $\tau_1 \in (0,1)$, let

$$\tau_2 = \tau_1 \beta(g_0) \sigma_{F_0,j}(\theta_*) / (E_{F_0}||m(W_i, \theta_*)||^2)^{1/2}. \qquad (14.30)$$

By (14.29), for all $g \in \mathcal{B}_{\rho_X}(g_0, \tau_2)$,

$$|m_j^*(g) - m_j^*(g_0)| \leq \tau_1 \beta(g_0) \text{ for all } j = 1, ..., k. \qquad (14.31)$$

Hence, for all $g \in \mathcal{B}_{\rho_X}(g_0, \tau_2)$, there exists $j \leq k$ such that either

$$\begin{aligned} & j \leq p \text{ and } m_j^*(g)/\beta(g_0) \leq m_j^*(g_0)/\beta(g_0) + \tau_1 = -1 + \tau_1 < 0 \text{ or} \\ & j \in \{p+1, ..., k\} \text{ and } |m_j^*(g)/\beta(g_0)| \geq |m_j^*(g_0)/\beta(g_0)| - \tau_1 = 1 - \tau_1 > 0 \end{aligned} \qquad (14.32)$$

using the triangle inequality.

By (14.32) and Assumption S3, $S(m^*(g)/\beta(g_0), h_{2,0}(g) + \varepsilon I_k) > 0$ for all $g \in \mathcal{B}_{\rho_X}(g_0, \tau_2)$. In addition, by Assumption Q, $Q(\mathcal{B}_{\rho_X}(g_0, \tau_2)) > 0$. These properties combine to give (14.28).

We use (14.28) in the following: for all $\delta > 0$,

$$
\begin{aligned}
& (n^{1/2}\beta(g_0))^{-\chi} T_n(\theta_*) \\
= \; & (n^{1/2}\beta(g_0))^{-\chi} \int_{\mathcal{G}} S\left(\nu_{n,F_0}(\theta_*, g) + h_{1,n,F_0}(\theta_*, g), \overline{h}_{2,n,F_0}(\theta_*, g)\right) dQ(g) \\
\geq \; & (n^{1/2}\beta(g_0))^{-\chi} \int_{\mathcal{B}_{\rho_X}(g_0,\tau_2)} S\left(\nu_{n,F_0}(\theta_*, g) + h_{1,n,F_0}(\theta_*, g), \overline{h}_{2,n,F_0}(\theta_*, g)\right) dQ(g) \\
= \; & \int_{\mathcal{B}_{\rho_X}(g_0,\tau_2)} S\left((n^{1/2}\beta(g_0))^{-1}\nu_{n,F_0}(\theta_*, g) + m^*(g)/\beta(g_0), \overline{h}_{2,n,F_0}(\theta_*, g)\right) dQ(g) \\
\rightarrow_p \; & \int_{\mathcal{B}_{\rho_X}(g_0,\tau_2)} S\left(m^*(g)/\beta(g_0),\, h_{2,0}(g) + \varepsilon I_k\right) dQ(g) \\
> \; & 0,
\end{aligned} \tag{14.33}
$$

where $\chi$ is as in Assumption S4, the first equality holds by (5.4), the first inequality holds by Assumption S1(c), the second equality holds by Assumption S4 and the definition of $m_j^*(g)$ in (6.2), the last inequality holds by (14.28), and the convergence holds by the argument given in the following paragraph.

By Lemma A1(a) and the continuous mapping theorem, $\sup_{g \in \mathcal{G}} ||\nu_{n,F_0}(\theta_*, g)|| = O_p(1)$. (Note that Lemma A1 applies for $(\theta_{a_n}, F_{a_n}) = (\theta_*, F_0) \notin \mathcal{F}$ for all $n \geq 1$ because Assumptions FA(b)-(d) imply conditions (ii)-(v) in the definition of $SubSeq(h_{2,F_0}(\theta_*))$.) Also, $(n^{1/2}\beta(g_0))^{-1} = o(1)$, because Assumptions FA and CI imply that $\beta(g_0) > 0$ for some $g_0 \in \mathcal{G}$. Hence, (i) $(n^{1/2}\beta(g_0))^{-1}\nu_{n,F_0}(\theta_*, \cdot) \Rightarrow 0_{\mathcal{G}}$. In addition, (ii) $\sup_{g \in \mathcal{G}} ||\overline{h}_{2,n,F_0}(\theta_*, g) - h_{2,0}(g) - \varepsilon I_k|| \rightarrow_p 0$, where $h_{2,0}(g) = h_{2,F_0}(\theta_*, g, g)$, by Lemma A1(b), (12.26), and the definition of $\overline{h}_{2,n,F}(\theta, g)$. As in previous proofs, by the almost sure representation theorem, there exists a probability space and random quantities defined on it with the same distributions as $(n^{1/2}\beta(g_0))^{-1}\nu_{n,F_0}(\theta_*, \cdot)$ and $\overline{h}_{2,n,F_0}(\theta_*, \cdot)$ for $n \geq 1$ such that the convergence in (i) and (ii) holds almost surely for these random quantities. Furthermore, using Assumptions S1(b) and S1(e), the integrand in the last equality in (14.33) is bounded by $\sup_{g \in \mathcal{B}^{cl}_{\rho_X}(g_0,\tau_2), \nu \in R^k : ||\nu|| \leq \delta_*} S(\nu + m^*(g)/\beta(g_0), (\varepsilon - \delta_{**})I_k) < \infty$ for all $g \in \mathcal{B}_{\rho_X}(g_0, \tau_2)$ for some $\delta_*, \delta_{**} > 0$ for $n$ sufficiently large, where $\mathcal{B}^{cl}_{\rho_X}(g_0, \tau_2)$ denotes the closure of $\mathcal{B}_{\rho_X}(g_0, \tau_2)$, because a continuous function on a compact set attains its supremum using Assumption S1(d) and using an argument analogous to that in (12.14) to treat the second argument of the function $S$. Thus, by the bounded convergence theorem, the convergence in (14.33) holds a.s. for the newly constructed random quantities. In consequence, it holds in probability for the original random quantities by the equality in distribution of the original and newly constructed random quantities. This completes the proof of the convergence in (14.33).

Next, we show that under $F_0$,

$$c(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_*), 1-\alpha) = O_p(1). \tag{14.34}$$

This and (14.33) give

$$\begin{aligned}
&P_{F_0}(T_n(\theta_*) > c(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_*), 1-\alpha)) \\
&= P_{F_0}\left((n^{1/2}\beta(g_0))^{-\chi} T_n(\theta_*) > (n^{1/2}\beta(g_0))^{-\chi} c(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_*), 1-\alpha)\right) \\
&\geq P_{F_0}\left(\int_{\mathcal{B}_{\rho_X}(g_0,\tau_2)} S\left(m^*(g)/\beta(g_0),\ h_{2,0}(g) + \varepsilon I_k\right) dQ(g) + o_p(1) > o_p(1)\right) \\
&\to 1
\end{aligned} \tag{14.35}$$

as $n \to \infty$, which establishes the result of the Theorem.

It remains to show (14.34). Lemma A5, applied with $h_{2,n} = h_{2,0}$, $\{h^*_{2,n} : n \geq 1\}$ being any sequence of $k \times k$-matrix-valued covariance kernels on $\mathcal{G} \times \mathcal{G}$ such that $d(h^*_{2,n}, h_{2,0}) \to 0$, $h_1 = 0_{\mathcal{G}}$, $\eta$ as in the definition of $c(h, 1-\alpha)$, $\alpha$ replaced by $\alpha - \eta > 0$, and $\eta_1 = \delta$, gives: $\forall \delta > 0$,

$$\begin{aligned}
&\liminf_{n\to\infty} \left[c_0(0_{\mathcal{G}}, h_{2,0}, 1-\alpha+\eta+\delta) + \delta - c_0(0_{\mathcal{G}}, h^*_{2,n}, 1-\alpha+\eta)\right] \geq 0 \text{ and hence} \\
&\limsup_{n\to\infty} c_0(0_{\mathcal{G}}, h^*_{2,n}, 1-\alpha+\eta) \leq c_0(0_{\mathcal{G}}, h_{2,0}, 1-\alpha+\eta+\delta) + \delta < \infty.
\end{aligned} \tag{14.36}$$

By Lemma A1(b) and (12.26), we obtain $d(\widehat{h}_{2,n}(\theta_*), h_{2,0}) \to_p 0$. As in previous proofs, by the almost sure representation theorem, there exists a probability space and random quantities $\tilde{h}_{2,n}(\cdot,\cdot)$ defined on it with the same distributions as $\widehat{h}_{2,n}(\theta_*,\cdot,\cdot)$ for $n \geq 1$ such that $d(\tilde{h}_{2,n}, h_{2,0}) \to 0$ a.s. This and (14.36) gives $\limsup_{n\to\infty} c_0(0_{\mathcal{G}}, \tilde{h}_{2,n}, 1-\alpha+\eta) < \infty$ a.s., which implies (14.34) because $\tilde{h}_{2,n}(\cdot,\cdot)$ and $\widehat{h}_{2,n}(\theta_*,\cdot,\cdot)$ have the same distribution for all $n \geq 1$ and $c(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_*), 1-\alpha) = c_0(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_*), 1-\alpha+\eta) + \eta$. $\square$

## 14.3 Proofs of Results for $\mathbf{n^{-1/2}}$-Local Alternatives

**Proof of Theorem 4.** The proof of part (a) uses the following. By element-by-element mean-value expansions about $\theta_n$ and Assumptions LA1(a), LA1(b), and LA2,

$$\begin{aligned} & D_{F_n}^{-1/2}(\theta_{n,*})E_{F_n}m(W_i,\theta_{n,*},g) \\ & = D_{F_n}^{-1/2}(\theta_n)E_{F_n}m(W_i,\theta_n,g)+\Pi_{F_n}(\theta_{n,g},g)(\theta_{n,*}-\theta_n), \text{ and so} \\ & n^{1/2}D_{F_n}^{-1/2}(\theta_{n,*})E_{F_n}m(W_i,\theta_{n,*},g)\rightarrow h_1(g)+\Pi_0(g)\lambda, \end{aligned} \tag{14.37}$$

where $\theta_{n,g}$ may differ across rows of $\Pi_{F_n}(\theta_{n,g},g)$, $\theta_{n,g}$ lies between $\theta_{n,*}$ and $\theta_n$, $\theta_{n,g}\rightarrow\theta_0$, $\Pi_{F_n}(\theta_{n,g},g)\rightarrow\Pi_0(g)$, and by definition $h_1(g)+\Pi_0(g)\lambda=\infty$ if $h_1(g)=\infty$.

Now, the proof of part (a) is the same as the proof of Theorem 2(b) with the following changes: (i) $(\theta_{n,*},F_n)$ appears in place of $(\theta_c,F_c)$ whenever $(\theta_c,F_c)$ is used in an expression with $n$ finite, (ii) $(\theta_0,F_0)$ appears in place of $(\theta_c,F_c)$ whenever $(\theta_c,F_c)$ is used in an asymptotic expression, (iii) $\{(\theta_{n,*},F_n):n\geq 1\}$ satisfies the conditions to be in $SubSeq(h_2)$ (where $h_2=h_{2,F_0}(\theta_0)$) by Assumptions LA1(a) and LA1(c)-(e) and because $\{W_i:i\geq 1\}$ are i.i.d. under $F_n$ and Assumption M holds given that $(\theta_n,F_n)\in\mathcal{F}$ by Assumption LA1, (iv) equation (14.16) is replaced by

$$\kappa_n^{-1}\overline{D}_{F_n}^{-1/2}(\theta_{n,*},g)D_{F_n}^{1/2}(\theta_{n,*})h_{1,n,F_n}(\theta_{n,*},g)\rightarrow\pi_1(g) \text{ as } n\rightarrow\infty, \tag{14.38}$$

which holds by Assumption LA4, (14.37) (because $\kappa_n^{-1}n^{1/2}\Pi_{F_n}(\theta_{n,g},g)(\theta_{n,*}-\theta_n)\rightarrow 0$), and $\overline{D}_{F_n}^{-1/2}(\theta_{n,*},g)\overline{D}_{F_n}^{1/2}(\theta_n,g)\rightarrow I_k$ (using Assumption LA1(c)), (v) $\pi_1(g)$ appears in place of $h_{1,\infty,F_c}(\theta_c,g)$ in (14.17), (vi) $\varphi(\pi_1(g))$ appears in place of $h_{1,\infty,F_c}(\theta_c,g)$ in (14.18)-(14.20), where (14.18) holds for all $g\in\mathcal{G}_\varphi$ by Assumption LA5(a) and (14.19) holds for all $g\in\mathcal{G}_\varphi$, (vii) Assumption LA5(b) is used in place of Assumption GMS2(a) in two places, (viii) $(h_1+\Pi_0\lambda,h_2)$ and $h_1(g)$ appear in place of $h_{\infty,F_c}(\theta_c)$ and $h_{1,\infty,F_c}(\theta_c)$, respectively, in (14.23) and (14.24), and (ix) (14.23) holds using (14.37) in place of $h_{1,n,F_c}(\theta_c)\rightarrow h_{1,\infty,F_c}(\theta_c)$ and using $\nu_{n,F_n}(\theta_{n,*},\cdot)\Rightarrow\nu_{h_2}(\cdot)$ in place of $\nu_{n,F_c}(\theta_c,\cdot)\Rightarrow\nu_{h_{2,F_c}(\theta_c)}(\cdot)$. The result $\nu_{n,F_n}(\theta_{n,*},\cdot)\Rightarrow\nu_{h_2}(\cdot)$ holds by Lemma A1(a) because $\{(\theta_{n,*},F_n):n\geq 1\}\in SubSeq(h_2)$ by the argument given in (iii) above. The desired result is given by (14.24) with the changes indicated above. This completes the proof of part (a).

Part (b) holds by the same argument as for part (a) but with $\varphi_n(\theta_{n,*},g)$ replaced by 0, which simplifies the argument considerably. Assumption LA6 is used in place of Assumption LA5(b) in the proof.

Part (c) holds by the following argument:

$$
\begin{aligned}
&\beta^{-\chi}T(h_1+\Pi_0\lambda_0\beta,h_2)\\
&=\beta^{-\chi}\int S(\nu_{h_2}(g)+h_1(g)+\Pi_0(g)\lambda_0\beta,h_2(g)+\varepsilon I_k)dQ(g)\\
&=\int S(\nu_{h_2}(g)/\beta+h_1(g)/\beta+\Pi_0(g)\lambda_0,h_2(g)+\varepsilon I_k)dQ(g)\\
&\to\int S(\Pi_0(g)\lambda_0,h_2(g)+\varepsilon I_k)dQ(g)>0 \qquad (14.39)
\end{aligned}
$$

as $\beta\to\infty$ a.s., where $\chi$ is as in Assumption S4, the second equality holds by Assumption S4, the convergence holds a.s. (with respect to the randomness in $\nu_{h_2}$) by the bounded convergence theorem applied for each fixed sample path $\omega$ because $||\nu_{h_2}(g)||$ has bounded sample paths a.s. and using Assumption LA3$'$ (which guarantees that $h_{1,j}(g)<\infty$ and hence $h_{1,j}(g)/\beta\to 0$ as $\beta\to\infty$ for all $j\leq p$, for all $g$ in a set with $Q$ measure one), and the inequality holds by Assumptions LA3$'$ and S3.

Equation (14.39) implies that $T(h_1+\Pi_0\lambda_0\beta,h_2)\to\infty$ a.s. as $\beta\to\infty$. Because $T(h_1+\Pi_0\lambda_0\beta,h_2)\sim J_{h,\beta\lambda_0}$ and the quantities $c(\varphi(\pi_1),h_2,1-\alpha)$ and $c(0_{\mathcal{G}},h_2,1-\alpha)$ do not depend on $\beta$, the result of part (c) follows. $\square$

## 14.4 Proofs Concerning the Verification of Assumptions S1-S4

**Proof of Lemma 1.** Assumptions S1(a)-(d) and S3 hold for the functions $S_1$, $S_2$, and $S_3$ by Lemma 1 of Andrews and Guggenberger (2009). Assumptions S1(e) and S4 hold immediately for the functions $S_1$, $S_2$, and $S_3$ with $\chi=2$ in Assumption S4.

To verify Assumption S2 for $S=S_1$, $S_2$, or $S_3$, it suffices to show that

$$\limsup_{n\to\infty}|S(m_n+\mu_n,\Sigma_n)-S(m_0+\mu_n,\Sigma_0)|=0 \qquad (14.40)$$

for all sequences $\{\mu_n\in[0,\infty)^p\times\{0\}^v:n\geq 1\}$ and $\{(m_n,\Sigma_n):n\geq 1\}$ such that $(m_n,\Sigma_n)\to(m_0,\Sigma_0)$, $m_0\in R^k$, and $\Sigma_0\in\mathcal{W}$.

For clarity of the proof, we consider a simple case first. We consider the function $S_1$ and suppose $\Sigma_n=\Sigma_0$. In this case, without loss of generality, we can assume that $\Sigma_0=I_k$. Given that $S_1$ is additive, it suffices to consider the cases where $(p,v)=(1,0)$ and $(0,1)$. It is easy to see that Assumption S2 holds in the latter case because $\mu_n$ does not appear. For the case where $(p, v) = (1, 0)$, we have

$$
\begin{aligned}
& |S_1(m_n + \mu_n, I_k) - S_1(m_0 + \mu_n, I_k)| \\
= & \left|\left([m_n + \mu_n]_-^2 - [m_0 + \mu_n]_-^2\right)\right| \\
\leq & \left|[m_n + \mu_n]_- - [m_0 + \mu_n]_-\right| \left([m_n + \mu_n]_- + [m_0 + \mu_n]_-\right) \\
\leq & |m_n - m_0| \left(|m_n| + |m_0|\right) \\
= & o(1)O(1), \qquad (14.41)
\end{aligned}
$$

where the second inequality holds because $\left|[a]_- - [b]_-\right| \leq |a - b|$ and by Assumption S1(b). This completes the verification of Assumption S2 for the simple case.

Next, we verify Assumption S2 for $S = S_2$. For any sequence $\{\mu_n \in [0, \infty)^p \times \{0\}^v : n \geq 1\}$, there exists a subsequence $\{u_n : n \geq 1\}$ of $\{n\}$ such that

$$
\begin{aligned}
& \lim_{n\to\infty} \left|S_2(m_{u_n} + \mu_{u_n}, \Sigma_{u_n}) - S_2(m_0 + \mu_{u_n}, \Sigma_0)\right| \\
= & \limsup_{n\to\infty} \left|S_2(m_n + \mu_n, \Sigma_n) - S_2(m_0 + \mu_n, \Sigma_0)\right|. \qquad (14.42)
\end{aligned}
$$

Let $\{t_{1,u_n}, t_{0,u_n} \in [0, \infty)^p \times \{0\}^v : n \geq 1\}$ be sequences such that

$$
\begin{aligned}
(m_{u_n} + \mu_{u_n} - t_{1,u_n})'\Sigma_{u_n}^{-1}(m_{u_n} + \mu_{u_n} - t_{1,u_n}) &\leq S_2(m_{u_n} + \mu_{u_n}, \Sigma_{u_n}) + 2^{-u_n} \text{ and} \\
(m_0 + \mu_{u_n} - t_{0,u_n})'\Sigma_0^{-1}(m_0 + \mu_{u_n} - t_{0,u_n}) &\leq S_2(m_0 + \mu_{u_n}, \Sigma_0) + 2^{-u_n}. \qquad (14.43)
\end{aligned}
$$

Then,

$$
\begin{aligned}
& \lim_{n\to\infty} [S_2(m_{u_n} + \mu_{u_n}, \Sigma_{u_n}) - S_2(m_0 + \mu_{u_n}, \Sigma_0)] \\
= & \lim_{n\to\infty} \left[(m_{u_n} + \mu_{u_n} - t_{1,u_n})'\Sigma_{u_n}^{-1}(m_{u_n} + \mu_{u_n} - t_{1,u_n}) - S_2(m_0 + \mu_{u_n}, \Sigma_0)\right] \\
\geq & \lim_{n\to\infty} [(m_{u_n} + \mu_{u_n} - t_{1,u_n})'\Sigma_{u_n}^{-1}(m_{u_n} + \mu_{u_n} - t_{1,u_n}) \\
& -(m_0 + \mu_{u_n} - t_{1,u_n})'\Sigma_0^{-1}(m_0 + \mu_{u_n} - t_{1,u_n})] \\
= & \lim_{n\to\infty} [(m_{u_n} + \mu_{u_n} - t_{1,u_n})'(\Sigma_{u_n}^{-1} - \Sigma_0^{-1})(m_{u_n} + \mu_{u_n} - t_{1,u_n}) \\
& +(m_{u_n} - m_0)'\Sigma_0^{-1}(m_0 + m_{u_n} + 2\mu_{u_n} - 2t_{1,u_n})] \\
= & 0, \qquad (14.44)
\end{aligned}
$$

where the last equality holds if $\mu_{u_n} - t_{1,u_n} = O(1)$ because $m_{u_n} \to m_0 < \infty$ and $\Sigma_{u_n}^{-1} \to \Sigma_0^{-1}$ as $n \to \infty$.

We now show that $\mu_{u_n} - t_{1,u_n} = O(1)$. We have

$$\begin{aligned} m_{u_n}' \Sigma_{u_n}^{-1} m_{u_n} &\geq S_2(m_{u_n} + \mu_{u_n}, \Sigma_{u_n}) \\ &\geq (m_{u_n} + \mu_{u_n} - t_{1,u_n})' \Sigma_{u_n}^{-1} (m_{u_n} + \mu_{u_n} - t_{1,u_n}) - 2^{-u_n}. \end{aligned} \tag{14.45}$$

Thus,

$$\begin{aligned} &\lim_{n\to\infty} (m_{u_n} + \mu_{u_n} - t_{1,u_n})' \Sigma_{u_n}^{-1} (m_{u_n} + \mu_{u_n} - t_{1,u_n}) \\ &\leq \lim_{n\to\infty} [m_{u_n}' \Sigma_{u_n}^{-1} m_{u_n} + 2^{-u_n}] = m_0' \Sigma_0^{-1} m_0 < \infty, \end{aligned} \tag{14.46}$$

which implies that $m_{u_n} + \mu_{u_n} - t_{1,u_n} = O(1)$. The latter and $m_{u_n} \to m_0 < \infty$ give

$$\mu_{u_n} - t_{1,u_n} = O(1). \tag{14.47}$$

Next, by an analogous argument to (14.44) with $\geq$ and $t_{1,u_n}$ replaced by $\leq$ and $t_{0,u_n}$, respectively, we obtain the following upper bound:

$$\begin{aligned} &\lim_{n\to\infty} [S(m_{u_n} + \mu_{u_n}, \Sigma_{u_n}) - S(m_0 + \mu_{u_n}, \Sigma_0)] \\ &= \lim_{n\to\infty} [S(m_{u_n} + \mu_{u_n}, \Sigma_{u_n}) - (m_0 + \mu_{u_n} - t_{0,n})' \Sigma_0^{-1} (m_0 + \mu_{u_n} - t_{0,u_n})] \\ &\leq 0, \end{aligned} \tag{14.48}$$

where the inequality uses $\mu_{u_n} - t_{0,u_n} = O(1)$, which holds by an analogous argument to that given for (14.47). Equations (14.44) and (14.48) imply that the left-hand side of (14.42) equals zero, which completes the verification of Assumption S2 for $S_2$.

The verification of Assumption S2 for $S = S_1$, where $\Sigma_n$ need not equal $\Sigma_0$, is obtained by replacing $\Sigma_n$ and $\Sigma_0$ in the proof above for $S_2$ by $Diag\{\Sigma_n\}$ and $Diag\{\Sigma_0\}$, respectively, because $S_1(m, \Sigma) = S_2(m, \Sigma)$ when $\Sigma$ is diagonal. Assumption S2 holds for the function $S_3$ when $(p, v) = (1, 0)$ and $(0, 1)$ because $S_3 = S_1 = S_2$ in these cases. It holds for $S_3$ in the general $(p, v)$ case because it holds in these two special cases. $\square$

# 15 Supplemental Appendix D

In this Appendix, we provide proofs of the results stated in Supplemental Appendix B. The first sub-section gives proofs for the Kolmogorov-Smirnov and approximate CvM tests and CS's. The second sub-section gives proofs for results concerning $\mathcal{G}_{B-spline}$ and $\mathcal{G}_{c/d}$. The third sub-section gives proofs for results concerning "asymptotic issues with the Kolmogorov-Smirnov statistic." The fourth sub-section gives proofs for the subsampling results.

## 15.1 Proofs of Kolmogorov-Smirnov and Approximate Cramér von Mises Results

**Proof of Lemma B1.** To verify Assumption S2$'$ for $S_1$, $S_2$, and $S_3$, it suffices to show that

$$\limsup_{n\to\infty} |S(m_n + \mu_n, \Sigma_n) - S(m_{n,0} + \mu_n, \Sigma_{n,0})| = 0 \tag{15.1}$$

for all sequences $\{\mu_n \in [0,\infty)^p \times \{0\}^v : n \geq 1\}$, $\{(m_n, \Sigma_n) \in \mathcal{M} \times \mathcal{W}_{bd} : n \geq 1\}$, and $\{(m_{n,0}, \Sigma_{n,0}) \in \mathcal{M} \times \mathcal{W}_{bd} : n \geq 1\}$ for which $(m_n, \Sigma_n) - (m_{n,0}, \Sigma_{n,0}) \to 0$ as $n \to \infty$.

The verification of (15.1) is an extension of the verification of (14.40) in the proof of Lemma 1. The extension consists of (i) replacing $m_0$ and $\Sigma_0$ by $m_{u_n,0}$ and $\Sigma_{u_n,0}$ throughout (14.42)-(14.48), (ii) making use of the fact that $m_{u_n}$, $m_{u_n,0}$, and $\Sigma_{u_n}^{-1}$ are bounded by the definitions of $\mathcal{M}$ and $\mathcal{W}_{bd}$, and (iii) making use of the fact that $\Sigma_{u_n}^{-1} - \Sigma_{u_n,0}^{-1} \to 0$ given that $\Sigma_{u_n} - \Sigma_{u_n,0} \to 0$ and $\Sigma_{u_n}, \Sigma_{u_n,0} \in \mathcal{W}_{bd}$. $\square$

**Proof of Theorem B1.** When $T_n(\theta)$ is the KS statistic and when $T_n(\theta)$ is replaced by the approximate statistic $\overline{T}_{n,s_n}(\theta)$, the results of Theorem 1 hold under the assumptions of that Theorem plus Assumption S2$'$. The proof of Theorem 1 goes through with the following changes: (i) the statistics $\tilde{T}_{a_n}$ and $\tilde{T}_{a_n,0}$ are changed from integrals with respect to $Q$ to suprema over $g \in \mathcal{G}_n$ or weighted averages over $\{g_1, ..., g_{s_n}\}$ with weights $\{w_{Q,n}(\ell) : \ell = 1, ..., s_n\}$, (ii) in the proof of (12.7), (12.10) holds uniformly over $g \in \mathcal{G}$ because Assumption S2 has been strengthened to Assumption S2$'$, and (iii) (12.11) holds with the supremum over $g \in \mathcal{G}_n$ added or with the average over $\{g_1, ..., g_{s_n}\}$ added, because (12.10) holds uniformly over $g \in \mathcal{G}$ and the weights are non-negative and sum to at most one by Assumption A2. This completes the proof of Theorem 1 for the KS and A-CvM test statistics.

The result of Theorem B1 is the same as that of Theorem 2(a). The proof of Theorem 2(a) follows immediately from Lemmas A2-A4. The proof of Lemma A4 uses Lemma A5. The proofs of Lemmas A2-A5 go through for the KS and A-CvM test statistics with the following minor changes: (i) in the proof of Lemma A2, $T(h)$ is replaced by $\overline{T}_{s_n}(h)$ (defined in (4.6)) and the new version of Theorem 1 for the KS and A-CvM statistics is employed, (ii) in the proof of Lemma A3, the form of the test statistic only enters through the first inequality of (12.23), which holds for the supremum and weighted average forms of the test statistic, (iii) in the proof of Lemma A4, no changes are required because the form of the test statistic only enters through Lemma A5, and (iv) in the proof of Lemma A5, $T(h)$ is replaced by $\overline{T}_{s_n}(h)$. $\square$

**Proof of Theorem B2.** Theorem B2 is proved by adjusting the proof Theorem 3. The proof of Theorem 3 goes through up to (14.32) with the only change being that $c(\cdot,\cdot,\cdot)$ is replaced by $c_{s_n}(\cdot,\cdot,\cdot)$ for A-CvM tests in (14.27)—in particular, the integral with respect to $Q$ in (14.28) is not changed. Equation (14.33) needs to be replaced, see (15.2) and (15.6) below; (14.34) is established with $c(\cdot,\cdot,\cdot)$ replaced by $c_{s_n}(\cdot,\cdot,\cdot)$ for A-CvM tests; (14.35) holds, with $T_n(\theta_*)$ and $c(\cdot,\cdot,\cdot)$ replaced by $\overline{T}_{n,s_n}(\theta_*)$ and $c_{s_n}(\cdot,\cdot,\cdot)$ for A-CvM tests, using the replacements for (14.33) given in (15.2) and (15.6) below; the first equation in (14.36) holds by Lemma A5 with $c(\cdot,\cdot,\cdot)$ replaced by $c_{s_n}(\cdot,\cdot,\cdot)$ for A-CvM tests, noting that Lemma A5 is extended to KS and A-CvM critical values in the proof of Theorem B1 above; in the second equation in (14.36) "$c_0(0_{\mathcal{G}}, h_{2,0}, 1-\alpha+\eta+\delta) < \infty$" holds for the KS critical value because $c_0(0_{\mathcal{G}}, h_{2,0}, 1-\alpha+\eta+\delta)$ does not depend on $n$ and the KS test statistic $T(0_{\mathcal{G}}, h_{2,0})$ is finite a.s. since the sample paths of $\nu_{h_{2,0}}(\cdot)$ and $h_{2,0}(\cdot)$ are bounded a.s.; and in the second equation in (14.36) "$\sup_{n\geq 1} c_{0,s_n}(0_{\mathcal{G}}, h_{2,0}, 1-\alpha+\eta+\delta) < \infty$" holds for an A-CvM critical value because $c_{0,s_n}(0_{\mathcal{G}}, h_{2,0}, 1-\alpha+\eta+\delta)$ is less than equal to the corresponding quantile based on the KS statistic, which does not depend on $n$ and is finite a.s.

For the KS test, we replace (14.33) with the following:

$$\begin{aligned}
& (n^{1/2}\beta(g_0))^{-\chi} T_n(\theta_*) \cdot Q(\mathcal{B}_{\rho_X}(g_0,\tau_2)) \\
= \ & (n^{1/2}\beta(g_0))^{-\chi} \sup_{g\in\mathcal{G}_n} S\left(\nu_{n,F_0}(\theta_*,g) + h_{1,n,F_0}(\theta_*,g), \overline{h}_{2,n,F_0}(\theta_*,g)\right) \cdot Q(\mathcal{B}_{\rho_X}(g_0,\tau_2)) \\
\geq \ & (n^{1/2}\beta(g_0))^{-\chi} \int_{\mathcal{B}_{\rho_X}(g_0,\tau_2)} 1(g\in\mathcal{G}_n) S\left(\nu_{n,F_0}(\theta_*,g) + h_{1,n,F_0}(\theta_*,g), \overline{h}_{2,n,F_0}(\theta_*,g)\right) dQ(g) \\
= \ & \int_{\mathcal{B}_{\rho_X}(g_0,\tau_2)} 1(g\in\mathcal{G}_n) S\left((n^{1/2}\beta(g_0))^{-1}\nu_{n,F_0}(\theta_*,g) + m^*(g)/\beta(g_0), \overline{h}_{2,n,F_0}(\theta_*,g)\right) dQ(g) \\
\to_p \ & \int_{\mathcal{B}_{\rho_X}(g_0,\tau_2)} S\left(m^*(g)/\beta(g_0),\ h_{2,0}(g) + \varepsilon I_k\right) dQ(g) > 0,
\end{aligned} \tag{15.2}$$

where $\chi$ is as in Assumption S4, $m^*(g) = (m_1^*(g), ..., m_k^*(g))'$, $m_j^*(g)$ is defined in (6.2) for $j \leq k$, $h_{2,0} = h_{2,F_0}(\theta_*)$, and the convergence uses the argument given in the paragraph following (14.33) as well as $1(g\in\mathcal{G}_n) \to 1(g\in\mathcal{G}) = 1$ as $n\to\infty$ by Assumption KS.

For A-CvM tests, we replace (14.33) with the following results:

$$\begin{aligned}
& (n^{1/2}\beta(g_0))^{-\chi} \overline{T}_{n,s_n}(\theta_*) \\
= & \sum_{\ell=1}^{s_n} w_{Q,n}(\ell) S\left((n^{1/2}\beta(g_0))^{-1}\nu_{n,F_0}(\theta_*,g_\ell) + m^*(g_\ell)/\beta(g_0), \overline{h}_{2,n,F_0}(\theta_*,g_\ell)\right),
\end{aligned} \tag{15.3}$$

using Assumption S4. We have

$$\sup_{g\in\mathcal{G}} |m_j^*(g)| \leq (E_{F_0}(m_j^2(W_i,\theta_*)/\sigma_{F_0,j}^2(\theta_*))^{1/2} (E_{F_0} G^2(X_i))^{1/2} < \infty, \tag{15.4}$$

for $j = 1, ..., k$, using the definition of $m^*(g)$, Assumption FA (which imposes Assumption M in part FA(e)), and the Cauchy-Schwarz inequality. Next, we have

$$\begin{aligned}
\sup_{g\in\mathcal{G}} \Big| & S\left((n^{1/2}\beta(g_0))^{-1}\nu_{n,F_0}(\theta_*,g) + m^*(g)/\beta(g_0), \overline{h}_{2,n,F_0}(\theta_*,g)\right) \\
& - S\left(m^*(g)/\beta(g_0), h_{2,0}(g) + \varepsilon I_k\right)\Big| = o_p(1)
\end{aligned} \tag{15.5}$$

under $F_0$, using the uniform continuity of $S$ over a compact set, which holds by Assumption S1(d), where attention can be restricted to a compact set by (i) equation (15.4), (ii) $\sup_{g\in\mathcal{G}} ||n^{-1/2}\nu_{n,F_0}(\theta_*,g)|| = o_p(1)$ by Lemma A1(a), and (iii) $\sup_{g\in\mathcal{G}} ||\overline{h}_{2,n,F_0}(\theta_*) - h_{2,0} - \varepsilon I_k|| = o_p(1)$ using Lemma A1(b) and the definition of $\overline{h}_{2,n,F_0}(\theta_*)$ in (5.2), and

Lemma A1 applies for the reasons given in the paragraph following (14.33).

Equations (15.3) and (15.5) yield

$$\begin{aligned}
&(n^{1/2}\beta(g_0))^{-\chi}\overline{T}_{n,s_n}(\theta_*) + o_p(1) \\
&= \sum_{\ell=1}^{s_n} w_{Q,n}(\ell) S\left(m^*(g_\ell)/\beta(g_0), h_{2,0}(g_\ell)\right) \\
&\to \int S\left(m^*(g)/\beta(g_0), h_{2,0}(g)\right) dQ(g) \\
&\geq \int_{\mathcal{B}_{\rho_X}(g_0,\tau_2)} S\left(m^*(g)/\beta(g_0), h_{2,0}(g)\right) dQ(g) > 0, \qquad (15.6)
\end{aligned}$$

where the convergence holds for fixed $\{g_1, g_2, ...\}$ by Assumptions A1, A2, and S4, the first inequality holds by Assumption S1(c), and the second inequality holds by (14.28). This completes the proof. $\square$

**Proof of Theorem B3.** Part (a) follows from part (b) because

$$c_{s_n}(\varphi_n(\theta_{n,*}), \widehat{h}_{2,n}(\theta_{n,*}), 1-\alpha) \leq c_{s_n}(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_{n,*}), 1-\alpha), \qquad (15.7)$$

which holds because $\varphi_n(\theta_*, g) \geq 0_k \; \forall g \in \mathcal{G}$ by Assumption GMS1(a), $c(h_1, \widehat{h}_{2,n}(\theta_*), 1-\alpha)$ is non-increasing in the first $p$ elements of $h_1$ by Assumption S1(b), and the last $v$ elements of $\varphi_n(\theta_*, g)$ equal zero.

Now, we prove part (b). When $T_n(\theta)$ is replaced by the A-CvM statistic $\overline{T}_{n,s_n}(\theta_{n,*})$, the results of Theorem 1 hold under Assumptions M, S1, and S2$'$ with $(\theta, F)$ replaced by $(\theta_{n,*}, F_n)$, $\sup_{(\theta,F)\in\mathcal{F}: h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}$ deleted, $T_n(\theta)$, $T(h_{n,F}(\theta))$, and $x_{h_{n,F}(\theta)}$ replaced by $\overline{T}_{n,s_n}(\theta_{n,*})$, $\overline{T}_{s_n}(h_{n,F_n}(\theta_{n,*}))$ (defined in (4.6)), and $x_{h_{n,F_n}(\theta_{n,*})}$, respectively, where $x_{h_{n,F_n}(\theta_{n,*})} \in R$ is a constant that may depend on $(\theta_{n,*}, F_n)$ and $n$ through $h_{n,F_n}(\theta_{n,*})$. The adjustments needed to the proof of Theorem 1 are quite similar to those stated at the beginning of the proof of Theorem B1. In addition, the proof uses the fact that $\{(\theta_{n,*}, F_n) : n \geq 1\}$ satisfies the conditions to be in $SubSeq(h_2)$ (where $h_2 = h_{2,F_0}(\theta_0)$) by Assumptions LA1(a) and LA1(c)-(e) and because $\{W_i : i \geq 1\}$ are i.i.d. under $F_n$ and Assumption M holds given that $(\theta_n, F_n) \in \mathcal{F}$ by Assumption LA1. Because $\{(\theta_{n,*}, F_n) : n \geq 1\} \in SubSeq(h_2)$, Lemma A1 applies, which is used in (12.3). Also, $(h_{1,n,F}(\theta), h_{2,F}(\theta))$ is changed to $(h_{1,n,F_n}(\theta_{n,*}), h_{2,F_n}(\theta_{n,*}))$ throughout the proof of Theorem 1.

Next, using the mean-value expansion in (14.37) and the definition $h_{1,n,F}(\theta, g) =$

$n^{1/2}D_F^{-1/2}(\theta)E_F m(W_i,\theta,g)$, we have:

$$\begin{aligned}
&\sup_{g\in\mathcal{G}}||h_{1,n,F_n}(\theta_{n,*},g)-h_{1,n,F_n}(\theta_n,g)-\Pi_0(g)\lambda|| \\
=\ &\sup_{g\in\mathcal{G}}||\Pi_{F_n}(\theta_{n,g},g)n^{1/2}(\theta_{n,*}-\theta_n)-\Pi_0(g)\lambda|| \\
\leq\ &\sup_{g\in\mathcal{G}}\ \sup_{\theta\in\Theta:||\theta-\theta_0||\leq\delta_n}||\Pi_{F_n}(\theta,g)\lambda(1+o(1))-\Pi_0(g)\lambda|| \\
\to\ &0, \qquad (15.8)
\end{aligned}$$

where $\theta_{n,g}$ may differ across rows of $\Pi_{F_n}(\theta_{n,g},g)$, $\theta_{n,g}$ lies between $\theta_{n,*}$ and $\theta_n$, $\delta_n = ||\theta_{n,*}-\theta_n||+||\theta_n-\theta_0||\to 0$, the inequality holds using Assumption LA1(a), and the convergence to zero uses Assumption LA2$'$(b). (Note that the $(1+o(1))$ term in (15.8) requires the condition in Assumption LA2$'$(b) that $\sup_{g\in\mathcal{G}}||\Pi_0(g)\lambda||<\infty$.)

Equation (15.8) and Assumption LA2$'$(a) give: for all $B<\infty$,

$$\sup_{g\in\mathcal{G}:h_1(g)\leq B}||h_{1,n,F_n}(\theta_{n,*},g)-h_1(g)-\Pi_0(g)\lambda||\to 0. \qquad (15.9)$$

By Assumption LA1(c), $d(h_{2,F_n}(\theta_{n,*}),h_{2,F_0}(\theta_0))\to 0$. This implies that $\nu_{h_{2,F_n}(\theta_{n,*})}(\cdot)\Rightarrow\nu_{h_2}(\cdot)$, where $h_2=h_{2,F_0}(\theta_0)$. As in previous proofs, by the almost sure representation theorem, there exists a probability space and random quantities $\tilde{\nu}_n(\cdot)$ and $\tilde{\nu}(\cdot)$ defined on it with the same distributions as $\nu_{h_{2,F_n}(\theta_{n,*})}(\cdot)$ and $\nu_{h_2}(\cdot)$, respectively, for $n\geq 1$, such that $\sup_{g\in\mathcal{G}}||\tilde{\nu}_n(g)-\tilde{\nu}(g)||\to 0$ a.s. Hence, $\overline{T}_{s_n}(h_{n,F_n}(\theta_{n,*}))$ and $\overline{\overline{T}}_{s_n}(h_{n,F_n}(\theta_{n,*}))$ have the same distribution, where the latter is defined to be

$$\overline{\overline{T}}_{s_n}(h_{n,F_n}(\theta_{n,*}))=\sum_{\ell=1}^{s_n}w_{Q,n}(\ell)S(\tilde{\nu}_n(g_\ell)+h_{1,n,F_n}(\theta_{n,*},g_\ell),h_{2,F_n}(\theta_{n,*},g_\ell)+\varepsilon I_k). \qquad (15.10)$$

For all $\beta>0$, $B<\infty$, and $\lambda=\lambda_0\beta$, we have

$$\begin{aligned}
A_{1,n}(\beta,B)\ =\ &\sup_{g\in\mathcal{G}:h_1(g)\leq B}|S(\tilde{\nu}_n(g)/\beta+h_{1,n,F_n}(\theta_{n,*},g)/\beta,h_{2,F_n}(\theta_{n,*},g)+\varepsilon I_k) \\
&\qquad -S(\tilde{\nu}(g)/\beta+h_1(g)/\beta+\Pi_0(g)\lambda_0,h_2(g)+\varepsilon I_k)| \\
\to\ &0 \text{ as } n\to\infty \text{ a.s.} \qquad (15.11)
\end{aligned}$$

using Assumption S2$'$, (15.9), $\sup_{g\in\mathcal{G}}||\tilde{\nu}_n(g)-\tilde{\nu}(g)||\to 0$ a.s., $\sup_{g\in\mathcal{G}}||\tilde{\nu}(g)||<\infty$ a.s., and $d(h_{2,F_n}(\theta_{n,*}),h_2)\to 0$, where $h_2=h_{2,F_0}(\theta_0)$.

In addition, for all $B < \infty$, we have

$$\begin{aligned} A_2(\beta, B) = & \sup_{g \in \mathcal{G}: h_1(g) \leq B} |S(\tilde{\nu}(g)/\beta + h_1(g)/\beta + \Pi_0(g)\lambda_0, h_2(g) + \varepsilon I_k) \\ & \qquad - S(\Pi_0(g)\lambda_0, h_2(g) + \varepsilon I_k)| \\ & \to 0 \text{ as } \beta \to \infty \text{ a.s.} \end{aligned} \tag{15.12}$$

We use (15.11) and (15.12) to obtain: for all constants $B_c^* < \infty$ as in Assumption A3,

$$\begin{aligned} & \beta^{-\chi} \overline{\overline{T}}_{s_n}(h_{n,F_n}(\theta_{n,*})) \\ \geq & \sum_{\ell=1}^{s_n} w_{Q,n}(\ell) 1(h_1(g_\ell) \leq B_c^*) S(\tilde{\nu}_n(g_\ell)/\beta + h_{1,n,F_n}(\theta_{n,*}, g_\ell)/\beta, h_{2,F_n}(\theta_{n,*}, g_\ell) + \varepsilon I_k) \\ \geq & \sum_{\ell=1}^{s_n} w_{Q,n}(\ell) 1(h_1(g_\ell) \leq B_c^*) S(\Pi_0(g_\ell)\lambda_0, h_2(g_\ell) + \varepsilon I_k) - A_{1,n}(\beta, B_c^*) - A_2(\beta, B_c^*) \\ \to_{n \to \infty \text{ a.s.}} & \int 1(h_1(g) \leq B_c^*) S(\Pi_0(g)\lambda_0, h_2(g) + \varepsilon I_k) dQ(g) - A_2(\beta, B_c^*) \\ \to_{\beta \to \infty \text{ a.s.}} & \int 1(h_1(g) \leq B_c^*) S(\Pi_0(g)\lambda_0, h_2(g) + \varepsilon I_k) dQ(g), \end{aligned} \tag{15.13}$$

where the first inequality uses Assumptions S1(c) and S4, the second inequality holds by the definitions of $A_{1,n}(\beta, B_c^*)$ and $A_2(\beta, B_c^*)$, the first convergence result holds by (15.11) and Assumption A3, and the second convergence result holds by (15.12).

Let $c_{\sup,0}(0_\mathcal{G}, h_2^*, 1-\alpha)$ denote the $1-\alpha$ quantile of $T_{\sup}(0_\mathcal{G}, h_2^*) = \sup_{g \in \mathcal{G}} S(\nu_{h_2}(g), h_2^*(g) + \varepsilon I_k)$, where $h_2^*$ is some $k \times k$-matrix-valued covariance kernel on $\mathcal{G} \times \mathcal{G}$. Let $0_{\mathcal{G} \times \mathcal{G}}$ denote the $k \times k$-matrix-valued covariance kernel on $\mathcal{G} \times \mathcal{G}$ that equals the $k \times k$ zero matrix for all $(g, g^*) \in \mathcal{G} \times \mathcal{G}$. The A-PA critical value satisfies

$$\begin{aligned} c_{s_n}(0_\mathcal{G}, \widehat{h}_{2,n}(\theta_{n,*}), 1-\alpha) & \leq c_{\sup,0}(0_\mathcal{G}, \widehat{h}_{2,n}(\theta_{n,*}), 1-\alpha+\eta) + \eta \\ & \leq c_{\sup,0}(0_\mathcal{G}, 0_{\mathcal{G} \times \mathcal{G}}, 1-\alpha+\eta) + \eta \\ & < \infty, \end{aligned} \tag{15.14}$$

where the first inequality holds because a weighted average over $\{g_1, ..., g_{s_n}\}$ with non-negative weights that sum to one or less (by Assumption A2) is less than or equal to the corresponding supremum over $g \in \mathcal{G}$, which implies that $\overline{T}_{s_n}(0_\mathcal{G}, h_2^*) \leq T_{\sup}(0_\mathcal{G}, h_2^*)$ $\forall h_2^*$, the second inequality holds because $S(\nu_{h_2}(g), h_2^*(g) + \varepsilon I_k) \leq S(\nu_{h_2}(g), \varepsilon I_k)$ $\forall g \in \mathcal{G}$, for all covariance kernels $h_2^*$ by Assumption S1(e), which implies that $T_{\sup}(0_{\mathcal{G}}, h_2^*) \leq T_{\sup}(0_{\mathcal{G}}, 0_{\mathcal{G}\times\mathcal{G}})$ $\forall h_2^*$, and the last inequality holds because $\sup_{g\in\mathcal{G}} S(\nu_{h_2}(g), \varepsilon I_k) < \infty$ a.s., which holds by Assumption S2$'$ and $\sup_{g\in\mathcal{G}} ||\nu_{h_2}(g)|| < \infty$ a.s.

We now have: for all $B_c^*$ as in Assumption A3,

$$\begin{aligned}
&\lim_{\beta\to\infty} \liminf_{n\to\infty} P_{F_n}\left(\overline{T}_{s_n}(h_{n,F_n}(\theta_{n,*})) > c_{s_n}(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_{n,*}), 1-\alpha)\right) \\
&\geq \lim_{\beta\to\infty} \liminf_{n\to\infty} P\left(\beta^{-\chi}\overline{\widetilde{T}}_{s_n}(h_{n,F_n}(\theta_{n,*})) > \beta^{-\chi}c(0_{\mathcal{G}}, 0_{\mathcal{G}\times\mathcal{G}}, 1-\alpha+\eta) + \beta^{-\chi}\eta\right) \\
&\geq \lim_{\beta\to\infty} P\left(\int 1(h_1(g) \leq B_c^*) S(\Pi_0(g)\lambda_0, h_2(g) + \varepsilon I_k) dQ(g) - A_2(\beta, B_c^*)\right. \\
&\qquad\qquad \left. > \beta^{-\chi}c(0_{\mathcal{G}}, h_2, 1-\alpha+\eta) + \beta^{-\chi}\eta\right) \\
&= 1\left(\int 1(h_1(g) \leq B_c^*) S(\Pi_0(g)\lambda_0, h_2(g) + \varepsilon I_k) dQ(g) > 0\right), \qquad (15.15)
\end{aligned}$$

where the first inequality holds by (15.14) and the equality in distribution of $\overline{\widetilde{T}}_{s_n}(h_{n,F_n}(\theta_{n,*}))$ and $\overline{T}_{s_n}(h_{n,F_n}(\theta_{n,*}))$, the second inequality holds by (i) the first two inequalities in (15.13), (ii) the first convergence result in (15.13), and (iii) the bounded convergence theorem, and the last equality holds by the second convergence result of (15.13) and the bounded convergence theorem.

The left-hand side (lhs) in (15.15) does not depend on $B_c^*$. Hence, the lhs is greater than or equal to the limit as $c \to \infty$ of the right-hand side, which equals

$$1\left(\int 1(h_1(g) \leq \infty) S(\Pi_0(g)\lambda_0, h_2(g) + \varepsilon I_k) dQ(g) > 0\right) = 1 \qquad (15.16)$$

by the monotone convergence theorem and the assumption that $B_c^* \to \infty$ as $c \to \infty$, where the equality holds by Assumptions LA3$'$ and S3.

Lastly, we prove part (c) regarding KS tests and CS's. The proof is essentially the same as that for parts (a) and (b) with $\overline{T}_{n,s_n}(\theta_{n,*})$, $c_{s_n}(\cdot,\cdot,\cdot)$, $\sum_{\ell=1}^{s_n} w_{Q,n}(\ell)...$, and $\int ... dQ(g)$ replaced by the KS quantities $T_n(\theta_{n,*})$, $c(\cdot,\cdot,\cdot)$, $\sup_{g\in\mathcal{G}}$, and $\sup_{g\in\mathcal{G}} ...$, respectively (or with $\mathcal{G}_n$ in place of $\mathcal{G}$). $\square$

## 15.2 Proof of Lemma B2 Regarding $\mathcal{G}_{\mathbf{B-spline}}$, $\mathcal{G}_{\mathbf{box,dd}}$, and $\mathcal{G}_{\mathbf{c/d}}$

**Proof of Lemma B2.** First we verify Assumption CI for $\mathcal{G} = \mathcal{G}_{B-spline}$. Let $m_{j,F}(\theta, x) = E_F(m_j(W_i, \theta) | X_i = x)$. Write

$$\mathcal{X}_F(\theta) = \left( \bigcup_{j=1}^{p} \{x \in R^{d_x} : m_{j,F}(\theta, x) < 0\} \right) \bigcup \left( \bigcup_{j=p+1}^{k} \{x \in R^{d_x} : m_{j,F}(\theta, x) \neq 0\} \right). \tag{15.17}$$

If $P_F(X_i \in \mathcal{X}_F(\theta)) > 0$, then the probability that $X_i$ lies in one of the $k$ sets in (15.17) is positive. Suppose (without loss of generality) that $P_F(X_i \in \{x : m_{1,F}(\theta, x) < 0\}) > 0$. The set $\{x : m_{1,F}(\theta, x) < 0\}$ can be written as the union of disjoint non-degenerate hypercubes in $\mathcal{C}_{B-spline}$ (i.e., hypercubes with positive Lebesgue volumes) because continuity of $m_{1,F}(\theta, x)$ implies that if $m_{1,F}(\theta, x) < 0$ then $m_{1,F}(\theta, y) < 0$ for all $y$ in some hypercube that includes $x$. The number of such hypercubes is countable (because otherwise their union would have infinite volume). One of these hypercubes, call it $H$, must have positive $X_i$ probability. (Otherwise, the union of these hypercubes would have $X_i$ probability zero.)

In sum, we have $H \in \mathcal{C}_{B-spline}$, $P_F(X_i \in H) > 0$, and $m_{1,F}(\theta, x) < 0$ for all $x \in H$. In addition, the B-spline whose support is $H$ is positive on the interior of $H$. Thus, if $P_F(X_i \in int(H)) > 0$, we have $E_F m_1(W_i, \theta) B_H(X_i) < 0$, which establishes Assumption CI.

On the other hand, if $P_F(X_i \in int(H)) = 0$, then we must have $P_F(X_i \in H/int(H)) > 0$. Because $m_{1,F}(\theta, x)$ is a continuous function of $x$, there exists a finite number of hypercubes in $\mathcal{C}_{B-spline}$ whose interiors have union that includes $H/int(H)$ and for which $m_{1,F}(\theta, x) < 0$ for all $x$ in each hypercube. One of these hypercubes, say $H_1$, must have interior with positive probability because $P_F(X_i \in H/int(H)) > 0$. In sum, $H_1 \in \mathcal{C}_{B-spline}$, $P_F(X_i \in int(H_1)) > 0$, $m_{1,F}(\theta, x) < 0$ for all $x \in H_1$, and the B-spline $B_{H_1}(x)$ is positive for $x \in int(H_1)$. Hence, $E_F m_1(W_i, \theta) B_{H_1}(X_i) < 0$, which establishes Assumption CI.

Now we establish Assumption CI for $\mathcal{G}_{box,dd}$. The fact that Assumption CI holds for $\mathcal{G} = \mathcal{G}_{box}$ for all $\bar{r} \in (0, \infty]$ by Lemma 3 implies that Assumption CI holds for $\mathcal{G} = \mathcal{G}_{box,dd}$ for all $\bar{r} \in (0, \infty]$. The reason is as follows. Let $\mathcal{G}_{box}(\bar{r})$ and $\mathcal{G}_{box,dd}(\bar{r})$ denote $\mathcal{G}_{box}$ and $\mathcal{G}_{box,dd}$, respectively, when $\bar{r}$ is the upper bound on $r_u$ or $r_{1,u}$ and $r_{2,u}$. For any box $C_{x_0,r} \in \mathcal{G}_{box}(\bar{r})$, if $C_{x_0,r}$ captures some deviation from the model, i.e., $E_F m_j(W_i, \theta) 1(X_i \in C_{x_0,r}) < 0$ for some $j = 1, ..., p$ or $E_F m_j(W_i, \theta) 1(X_i \in C_{x_0,r}) \neq$

0 for some $j = p+1, ..., k$, then (i) $C_{x_0,r} \cap Supp_{F_{X,0}}(X_i) \neq \phi$ and (ii) $C_{x_0+\eta,r+\eta}$ captures the same deviation for $\eta > 0$ sufficiently small. Result (ii) holds because $\lim_{\eta \downarrow 0} E_F m_j(W_i,\theta) 1(X_i \in C_{x_0+\eta,r+\eta}) = E_F m_j(W_i,\theta)1(X_i \in C_{x_0,r})$. The latter holds by the bounded convergence theorem because $(C_{x_0+\eta,r+\eta} - C_{x_0,r}) \downarrow \phi$ as $\eta \downarrow 0$, and hence $m_j(w,\theta)1(x \in C_{x_0+\eta,r+\eta}) \to m_j(w,\theta)1(x \in C_{x_0,r})$ as $\eta \downarrow 0$ for every $w$, and $E_F|m_j(W_i,\theta)1(X_i \in C_{x_0+\eta,r+\eta})| \leq E_F|m_j(W_i,\theta)| < \infty$. By (i) and $\eta \in (0, \bar{r}/2]$, $C_{x_0+\eta,r+\eta}$ can be written as a box, $C_{x,r_1,r_2}$ in $\mathcal{G}_{box,dd}(3\bar{r})$ by picking a point $x \in C_{x_0,r} \cap Supp_{F_{X,0}}(X_i)$, which is necessarily in the interior of $C_{x_0+\eta,r+\eta}$, and letting $r_1 = x - x_0 + r$ and $r_2 = x_0 + r - x + 2\eta$. We have $|x - x_0| \leq \bar{r}$, $r_1 \leq 2\bar{r}$, and $r_2 \leq 3\bar{r}$. Because $C_{x,r_1,r_2} = C_{x_0+\eta,r+\eta}$ and $C_{x_0+\eta,r+\eta}$ captures a deviation from the model, $C_{x,r_1,r_2}$ does as well, and the proof is complete.

Note that in the preceding argument, it is necessary to expand $C_{x_0,r}$ to $C_{x_0+\eta,r+\eta}$ because $C_{x_0,r}$ is not necessarily in $\mathcal{G}_{box,dd}(3\bar{r})$ if the only elements of $C_{x_0,r} \cap Supp_{F_{X,0}}(X_i)$ are on the boundary of $C_{x_0,r}$. Also, note that the argument above does not go through if one uses symmetric side lengths (i.e., $r_{1,u} = r_{2,u}$) in the definition of $\mathcal{G}_{box,dd}$.

Next, we verify Assumption CI for $\mathcal{G} = \mathcal{G}_{c/d}$. We write

$$\begin{aligned}
\mathcal{X}_F(\theta) &= \cup_{d \in D} \mathcal{X}_{1,F}(\theta, d), \text{ where} \qquad (15.18)\\
\mathcal{X}_{1,F}(\theta,d) &= \{x_1 \in R^{d_{x,1}} : E_F(m_j(W_i,\theta)\,|X_{1,i} = x_1, X_{2,i} = d) < 0 \text{ for some } j \leq p \text{ or}\\
&\qquad E_F(m_j(W_i,\theta)\,|X_{1,i} = x_1, X_{2,i} = d) \neq 0 \text{ for some } j = p+1, ..., k\},
\end{aligned}$$

for $d \in D$. We have

$$\begin{aligned}
P_F(X_i \in \chi_F(\theta)) &= P_F\left((X_{1,i}', X_{2,i}')' \in \bigcup_{d \in D} \chi_{1,F}(\theta, d)\right)\\
&= \sum_{d \in D} P_F(X_{1,i} \in \chi_{1,F}(\theta,d)|X_{2,i} = d) P_F(X_{2,i} = d). \qquad (15.19)
\end{aligned}$$

If $P_F(X_i \in \chi_F(\theta)) > 0$, then there exists some $d^* \in D$ such that $P_F(X_{2,i} = d^*) > 0$ and

$$P_F((X_{1,i} \in \chi_{1,F}(\theta, d^*)|X_{2,i} = d^*) > 0. \qquad (15.20)$$

Given the inequality in (15.20), we use the same argument to verify Assumption CI as given for $\mathcal{G}_{c\text{-}cube}$, $\mathcal{G}_{box}$, $\mathcal{G}_{B-spline}$, or $\mathcal{G}_{box,dd}$ with $d_x$ replaced by $d_{x,1}$, but with $E_F(\cdot)$ replaced by $E_F(\cdot|X_{2,i} = d^*)$ throughout, and using the fact that $\{g : g = g_1 1_{\{d^*\}},$

$g_1 \in \mathcal{G}_1\} \subset \mathcal{G}_{c/d}$ for $\mathcal{G}_1 = \mathcal{G}_{c\text{-}cube}$, $\mathcal{G}_{box}$, $\mathcal{G}_{B-spline}$, or $\mathcal{G}_{box,dd}$.

Next, we verify Assumption M. Assumptions M(a) and M(b) hold for $\mathcal{G}_{B-spline}$ by taking $G(x) = 2/3$ $\forall x$ and $\delta_1 = 4/\delta + 3$. Assumption M(c) holds for $\mathcal{G}_{B-spline}$ because each element of $\mathcal{G}_{B-spline}$ can be written as the sum of four functions each of which is the product of an indicator function of a box and a polynomial of order four. Manageability of polynomials and indicator functions of boxes hold because they have finite pseudo-dimension (as defined in Pollard (1990, Sec. 4)). Manageability of finite linear combinations of these functions holds by the stability properties of cover numbers under addition and pointwise multiplication, see Pollard (1990, Sec. 5).

Assumption M holds for $\mathcal{G}_{box,dd}$ because it holds for $\mathcal{G}_{box}$ by Lemma 3 and $\mathcal{G}_{box,dd} \subset \mathcal{G}_{box}$.

The verification of Assumption M for $\mathcal{G} = \mathcal{G}_{c/d}$ is the same as in the proof of Lemma 3 when $\mathcal{G}_1$ is $\mathcal{G}_{c\text{-}cube}$, $\mathcal{G}_{box}$, or $\mathcal{G}_{box,dd}$ because $\mathcal{C}_{box} \times \{\{d\} : d \in D\}$ is a Vapnik-Cervonenkis class of sets. The verification of Assumption M for $\mathcal{G} = \mathcal{G}_{c/d}$ when $\mathcal{G}_1$ is $\mathcal{G}_{B-spline}$ is essentially the same as the proof above for $\mathcal{G}_{B-spline}$. The functions in $\mathcal{G}_{c/d}$ in this case still can be written as the sum of four functions each of which is the product of an indicator function of a box—in this case, the box is of the form $B \times \{d\}$, where $B$ is a box in $R^{d_{x,1}}$ and $d \in D$—and a polynomial of order four.

Assumption FA(e) holds for $\mathcal{G}_{B-spline}$, $\mathcal{G}_{box,dd}$, and $\mathcal{G}_{c/d}$ by the same arguments as given above for Assumption M.

This completes the proofs of parts (a)-(d) of the Lemma.

Part (e) of the Lemma holds, i.e., $Supp(Q_c) = \mathcal{G}_{B-spline}$, because $\mathcal{G}_{B-spline}$ is countable and $Q_c$ has a probability mass function that is positive at each element in $\mathcal{G}_{B-spline}$.

Now, we prove part (f) using a similar argument to that for part (b) of Lemma 4. Consider $g = g_{x,r_1,r_2} \in \mathcal{G}_{box,dd}$, where $g_{x,r_1,r_2}(y) = 1(y \in C_{x,r_1,r_2}) \cdot 1_k$ and $(x, r_1, r_2) \in Supp(X_i) \times (\times_{u=1}^{d_x}(0, \sigma_{X,u}\bar{r}))^2$. Let $\delta > 0$ be given. Let $\eta_0 = (\eta_{0,1}, ..., \eta_{0,d_x})'$ and likewise for $\eta_1$ and $\eta_2$. Define

$$G_{g,\bar{\eta}} = \{g_{x+\eta_0,r_1-\eta_1,r_2+\eta_2} : -\bar{\eta} \leq \eta_{0,u} \leq \bar{\eta},\ \bar{\eta} \leq \eta_{1,u}, \eta_{2,u} \leq 2\bar{\eta}\ \forall u \leq d_x\}. \qquad (15.21)$$

By the same sort of argument as for (14.26), for $g^* = g_{x+\eta_0,r_1-\eta_1,r_2+\eta_2} \in G_{g,\bar{\eta}}$, we have

$$
\begin{aligned}
\rho_X^2(g, g^*) &= E_{F_{X,0}}[1(X_i \in C_{x,r_1,r_2}) - 1(X_i \in C_{x+\eta_0,r_1-\eta_1,r_2+\eta_2})]^2 \\
&\leq \sum_{u=1}^{d_x}[P_{F_{X,0}}(X_{i,u} \in (x_u - r_{1,u}, x_u + \eta_{0,u} - (r_{1,u} - \eta_{1,u})]) \\
&\quad + P_{F_{X,0}}(X_{i,u} \in (x_u + r_{2,u}, x_u + \eta_{0,u} + r_{2,u} + \eta_{2,u}])] \\
&\leq \sum_{u=1}^{d_x}[F_{X_u,0}(x_u - r_{1,u} + 3\overline{\eta}) - F_{X_u,0}(x_u - r_{1,u})] \\
&\quad + \sum_{u=1}^{d_x}[F_{X_u,0}(x_u + r_{2,u} + 3\overline{\eta}) - F_{X_u,0}(x_u + r_{2,u})], \qquad (15.22)
\end{aligned}
$$

where $F_{X_u,0}(\cdot)$ denotes the distribution function of $X_{i,u}$ and the first inequality holds because $\eta_{0,u} + \eta_{1,u} \geq 0$ and $\eta_{0,u} + \eta_{2,u} \geq 0$. Because distribution functions are right continuous, the rhs of (15.22) converges to zero as $\overline{\eta} \downarrow 0$. Thus, $\rho_X^2(g, g^*)$ converges to zero uniformly over $G_{g,\overline{\eta}}$ as $\overline{\eta} \downarrow 0$ and there exists an $\overline{\eta} > 0$ sufficiently small that $G_{g,\overline{\eta}} \subset \mathcal{B}_{\rho_X}(g, \delta)$.

Next, we have $Q_c(G_{g,\overline{\eta}})$ equals

$$
Q^*_{F_{X,0}}\left(\times_{u=1}^{d_x}[x_u - \overline{\eta}, x_u + \overline{\eta}] \times_{u=1}^{d_x} [r_{1,u} - 2\overline{\eta}, r_{1,u} - \overline{\eta}] \times_{u=1}^{d_x} [r_{2,u} + \overline{\eta}, r_{2,u} + 2\overline{\eta}]\right) > 0, \qquad (15.23)
$$

where $Q^*_{F_{X,0}} = F_{X,0} \times Unif((\times_{u=1}^{d_x}(0, \sigma_{X,u}\overline{r}))^2$ and the inequality holds because $x \in Supp(X_i)$ and $\overline{\eta} > 0$. This completes the proof of part (f).

Lastly, we prove part (g). By parts (e) and (f) and parts (a) and (b) of Lemma 4, we have $\mathcal{G}_1 \subset Supp(Q_1)$. Because $Supp(Q_D) = D$ and $Q_e = Q_1 \times Q_D$, we have $\mathcal{G}_{c/d} \subset Supp(Q_e)$. $\square$

## 15.3 Proofs of Theorems B4 and B5 Regarding Uniformity Issues

**Proof of Theorem B4.** Part (a) holds by an empirical process central limit theorem because the intervals $\{(a, b] : 0 \leq a < b \leq 1\}$ form a Vapnik-Cervonenkis class of sets, e.g., see the proof of Lemma A1(a). The covariance kernel of $\nu(\cdot)$ and the pseudo-metric $\rho_*$ are specified below.

Let $c \vee d = \max\{c, d\}$ and $c \wedge d = \min\{c, d\}$.

To prove part (b), we write

$$
\begin{aligned}
Y_i g_{a,b}(X_i) &= (U_i + 1(X_i \in (\varepsilon_n, 1]) \cdot 1(X_i \in (a,b]) \\
&= U_i 1(X_i \in (a,b]) + 1(X_i \in (a \vee \varepsilon_n, b])
\end{aligned}
\tag{15.24}
$$

and

$$
\begin{aligned}
E_{F_n} Y_i g_{a,b}(X_i) &= E_{F_n} U_i 1(X_i \in (a,b]) + P_{F_n}(X_i \in (a \vee \varepsilon_n, b]) \\
&= P_{F_n}(X_i \in (a \vee \varepsilon_n, b]) \\
&\to (b-a)/2,
\end{aligned}
\tag{15.25}
$$

where the second equality uses Assumption CX(b) and the convergence uses Assumption CX(c) and holds by slightly different arguments when $a = 0$ and $a > 0$. Equation (15.25) and $b - a > 0$ imply that $h_{1,n}(g_{a,b}) = n^{1/2} E_{F_n} Y_i g_{a,b}(X_i) \to \infty = h_1(g_{a,b})$ as $n \to \infty$ for all $g_{a,b} \in \mathcal{G}$, which proves part (b).

Part (c) holds because $h_1(g_{a,b}) = \infty$ for all $g_{a,b} \in \mathcal{G}$ and

$$
\begin{aligned}
\inf_{g_{a,b} \in \mathcal{G}} h_{1,n}(g_{a,b}) &= \inf_{g_{a,b} \in \mathcal{G}} n^{1/2} P_{F_n}(X_i \in (a \vee \varepsilon_n, b]) \\
= \inf_{a,b:\ \varepsilon_n \le a < b \le 1} & n^{1/2} P_{F_n}(X_i \in (a,b]) = 0
\end{aligned}
\tag{15.26}
$$

for all $n$, where the first equality holds by (15.25) and the last equality holds by Assumption CX(c).

Part (d) holds because $\nu_n(g_{a,b}) + h_{1,n}(g_{a,b}) = O_p(1) + n^{1/2}(b-a)/2 \to_p \infty$ by part (a) and (15.25) for all $g_{a,b} \in \mathcal{G}$. This, combined with Assumption CX(f) (in particular, Assumption S1(d)), proves part (d).

Part (e) holds by part (b) and Assumption CX(f) (in particular, Assumption S2) because $S(\nu(g_{a,b}) + h_1(g_{a,b})) = S(\infty) = 0$ for all $g_{a,b} \in \mathcal{G}$.

To show part (f), we define

$$
g_n^*(x) = 1(x \in (0, \varepsilon_n]). \tag{15.27}
$$

Then,

$$
h_{1,n}(g_n^*) = n^{1/2} E_{F_n} Y_i g_n^*(X_i) = P_{F_n}(X_i \in (0 \vee \varepsilon_n, \varepsilon_n]) = 0 \tag{15.28}
$$

for all $n$, where the second equality holds by (15.25) with $a = 0$ and $b = \varepsilon_n$.

Next, we have

$$\sup_{g_{a,b}\in\mathcal{G}} S(\nu_n(g_{a,b}) + h_{1,n}(g_{a,b})) \geq S(\nu_n(g_n^*) + h_{1,n}(g_n^*)) = S(\nu_n(g_n^*)), \tag{15.29}$$

where the equality holds by (15.28). The asymptotic distribution of $S(\nu_n(g_n^*))$ is established as follows:

$$\begin{aligned}
\nu_n(g_n^*) \;&=\; n^{-1/2}\sum_{i=1}^n [Y_i 1(X_i \in (0,\varepsilon_n]) - E_{F_n} Y_i 1(X_i \in (0,\varepsilon_n])] \\
&=\; n^{-1/2}\sum_{i=1}^n [U_i 1(X_i = \varepsilon_n) + U_i 1(X_i \in (0,\varepsilon_n)) \\
&\qquad +1(X_i \in (\varepsilon_n, 1])1(X_i \in (0,\varepsilon_n]) - E_{F_n} 1(X_i \in (\varepsilon_n,1])1(X_i \in (0,\varepsilon_n])] \\
&=\; n^{-1/2}\sum_{i=1}^n U_i 1(X_i = \varepsilon_n) + n^{-1/2}\sum_{i=1}^n U_i 1(X_i \in (0,\varepsilon_n)) \\
&\to_d\; Z^* \sim N(0,1/2),
\end{aligned} \tag{15.30}$$

where the second equality uses $E_{F_n} U_i = 0$ and $U_i$ and $X_i$ are independent. The convergence in distribution in (15.30) holds by a triangular array CLT for the first summand on the second last line because $U_i 1(X_i = \varepsilon_n)$ has mean zero and variance $E_{F_n} U_i^2 1(X_i = \varepsilon_n) = 1 \cdot P_{F_n}(X_i = \varepsilon_n) = 1/2$ for all $n$ using Assumption CX(b). The second summand on the second last line of (15.30) is $o_p(1)$ because its mean is zero and its variance is

$$\begin{aligned}
&Var\left(n^{-1/2}\sum_{i=1}^n U_i 1(X_i \in (0,\varepsilon_n))\right) = Var(U_i 1(X_i \in (0,\varepsilon_n))) \\
&= E_{F_n} U_i^2 1(X_i \in (0,\varepsilon_n)) = 1 \cdot P_{F_n}(X_i \in (0,\varepsilon_n)) = \varepsilon_n/2,
\end{aligned} \tag{15.31}$$

where the first equality holds by Assumption CX(d), the second and third equalities hold by Assumption CX(b), and the last equality holds by Assumption CX(c).

Equations (15.29) and (15.30), Assumption S1(d), and the continuous mapping theorem combine to prove part (f).

Part (g) holds if

$$\sup_{g_{a,b}\in\mathcal{G}} S(\nu_n(g_{a,b}) + h_{1,n}(g_{a,b})) \nrightarrow_p 0 \tag{15.32}$$

using part (e). By part (f), for all $\delta \geq 0$,

$$\begin{aligned} \liminf_{n\to\infty} P\left(\sup_{g_{a,b}\in\mathcal{G}} S(\nu_n(g_{a,b}) + h_{1,n}(g_{a,b})) > \delta\right) &\geq \liminf_{n\to\infty} P\left(S(\nu_n(g_n^*)) > \delta\right) \\ &= P(S(Z^*) > \delta). \end{aligned} \tag{15.33}$$

Now, by the dominated convergence theorem, as $\delta \to 0$,

$$P(S(Z^*) > \delta) \to P(S(Z^*) > 0) = 1/2, \tag{15.34}$$

where the equality holds because $S(m) > 0$ iff $m < 0$ by Assumption S2 and $P(Z^* < 0) = 1/2$. Hence, the right-hand side in (15.33) is arbitrarily close to 1/2 for $\delta > 0$ sufficiently small, which implies that (15.32) holds and part (g) is established.

Lastly, we compute the covariance kernel $K(g_{a_1,b_1}, g_{a_2,b_2})$ of the Gaussian process $\nu(\cdot)$. We have

$$\begin{aligned} & E_{F_n} Y_i^2 g_{a_1,b_1}(X_i) g_{a_2,b_2}(X_i) \\ = \;& E_{F_n}(U_i + 1(X_i \in (\varepsilon_n, 1])^2 \cdot 1(X_i \in (a_1 \vee a_2, b_1 \wedge b_2]) \\ = \;& E_{F_n} U_i^2 1(X_i \in (a_1 \vee a_2, b_1 \wedge b_2]) \\ & + E_{F_n}(2U_i + 1) 1(X_i \in (a_1 \vee a_2 \vee \varepsilon_n, b_1 \wedge b_2]) \\ = \;& P_{F_n}(X_i \in (a_1 \vee a_2, b_1 \wedge b_2]) + P_{F_n}(X_i \in (a_1 \vee a_2 \vee \varepsilon_n, b_1 \wedge b_2]) \\ \to \;& (1/2) 1(a_1 = a_2 = 0) + \max\{(b_1 \wedge b_2) - (a_1 \vee a_2), 0\} \\ = \;& K_1(g_{a_1,b_1}, g_{a_2,b_2}), \end{aligned} \tag{15.35}$$

where the third equality uses Assumption CX(b) and the convergence uses Assumption CX(c).

In addition, we have

$$\lim_{n\to\infty} E_{F_n} Y_i g_{a,b}(X_i) = (b-a)/2 = K_2(g_{a,b}), \tag{15.36}$$

where the first equality holds by (15.25). Putting the results of (15.35) and (15.36) together yields

$$\begin{aligned}
&K(g_{a_1,b_1}, g_{a_2,b_2}) \\
&= \lim_{n\to\infty} \left( E_{F_n} Y_i^2 g_{a_1,b_1}(X_i) g_{a_2,b_2}(X_i) - E_{F_n} Y_i g_{a_1,b_1}(X_i) \cdot E_{F_n} Y_i g_{a_2,b_2}(X_i) \right) \\
&= K_1(g_{a_1,b_1}, g_{a_2,b_2}) - K_2(g_{a_1,b_1}) K_2(g_{a_2,b_2}).
\end{aligned} \tag{15.37}$$

The square of the pseudo-metric $\rho_*$ on $\mathcal{G}$ is

$$\begin{aligned}
&\rho_*^2(g_{a_1,b_1}, g_{a_2,b_2}) \\
&= \lim_{n\to\infty} E_{F_n} \left( Y_i g_{a_1,b_1}(X_i) - Y_i g_{a_2,b_2}(X_i) - E_{F_n} Y_i g_{a_1,b_1}(X_i) + E_{F_n} Y_i g_{a_2,b_2}(X_i) \right)^2.
\end{aligned} \tag{15.38}$$

The limit in (15.38) exists and can be computed via calculations analogous to those in (15.25) and (15.35). □

**Proof of Theorem B5.** For notational convenience, we let $g$ denote $g_{a,b}$. By Theorem B4(a), $\nu_n(\cdot) \Rightarrow \nu(\cdot)$ as $n \to \infty$. As in the proof of Theorem 1(a), by an almost sure representation argument, e.g., see Thm. 9.4 of Pollard (1990), there exist processes $\tilde{\nu}_n(\cdot)$ and $\tilde{\nu}(\cdot)$ on $\mathcal{G}$ that have the same distributions as $\nu_n(\cdot)$ and $\nu(\cdot)$, respectively, for which

$$\sup_{g\in\mathcal{G}} |\tilde{\nu}_n(g) - \tilde{\nu}(g)| \to 0 \text{ a.s.} \tag{15.39}$$

Let $\widetilde{\Omega}$ denote the sample paths for which the convergence in (15.39) holds. By (15.39), $P(\widetilde{\Omega}) = 1$.

For each $\omega \in \widetilde{\Omega}$, we apply the bounded convergence theorem to obtain

$$\lim_{n\to\infty} \int S(\tilde{\nu}_n(g)(\omega) + h_{1,n}(g)) dQ(g) = \int S(\tilde{\nu}(g)(\omega) + h_1(g)) dQ(g), \tag{15.40}$$

which yields the result of the Theorem. Now we check the conditions for the bounded convergence theorem. For all $g \in \mathcal{G}$, pointwise convergence holds:

$$S(\tilde{\nu}_n(g)(\omega) + h_{1,n}(g)) \to S(\tilde{\nu}(g)(\omega) + h_1(g)) \text{ as } n \to \infty$$

by (15.39), Theorem B4(b), and Assumption S1(d). A bound on $S(\tilde{\nu}_n(g)(\omega) + h_{1,n}(g))$ over $g \in \mathcal{G}$ and $n$ sufficiently large is given by $S(\inf_{g^*\in\mathcal{G}} \tilde{\nu}(g^*)(\omega) - \varepsilon)$ for some $\varepsilon > 0$.

This follows because for all $\varepsilon > 0$ and $g \in \mathcal{G}$, we have

$$\begin{aligned} 0 &\leq S(\tilde{\nu}_n(g)(\omega) + h_{1,n}(g)) \leq S(\tilde{\nu}_n(g)(\omega)) \\ &\leq S(\inf_{g^* \in \mathcal{G}} \tilde{\nu}_n(g^*)(\omega)) \leq S(\inf_{g^* \in \mathcal{G}} \tilde{\nu}(g^*)(\omega) - \varepsilon) < \infty, \end{aligned} \tag{15.41}$$

where the first inequality holds by Assumption S1(c), the second inequality holds by Assumption S1(b) and $h_{1,n}(g) \geq 0$ for all $g \in \mathcal{G}$ by (15.25), the third inequality holds by Assumption S1(b), the fourth inequality holds for all $n$ sufficiently large by (15.39) and Assumption S1(b), and the last inequality holds because $\inf_{g^* \in \mathcal{G}} \tilde{\nu}(g^*)(\omega) > -\infty$ because the sample paths of $\tilde{\nu}(\cdot)$ are bounded a.s. (which follows from $|m(W_i, \theta_0)g(X_i)| \leq |m(W_i, \theta_0)| \leq |U_i| + 1 < \infty$ a.s. and (15.39)). This completes the proof of (15.40) and the Theorem is proved. $\square$

## 15.4 Proofs of Subsampling Results

**Proof of Lemma B3.** For $S_1$, Assumption SQ(a) holds because (i) if $v \geq 1$, the summand $\sum_{j=p+1}^{k}(\nu_{h_2,j}^2(g)/(h_{2,j,j}(g) + \varepsilon))$ is absolutely continuous for all $g \in \mathcal{G}$, where $\nu_{h_2}(g) = (\nu_{h_2,1}(g), ..., \nu_{h_2,k}(g)'$ and $h_{2,j,j}(g)$ denotes the $j$th diagonal element of $h_2(g)$, (ii) if $v = 0$ and $h_1(g) \neq \infty^p$, the summands $[\nu_{h_2,j}(g) + h_{1,j}(g)]_-^2/(h_{2,j,j}(g) + \varepsilon)$ are absolutely continuous for $x > 0$ and all $j \leq p$ such that $h_{1,j}(g) < \infty$, (iii) if $v = 0$ and $h_1(g) = \infty^p$, $S_1(\nu_{h_2}(g) + h_1(g), h_2(g) + \varepsilon I_k) = 0$ and its distribution function equals one for all $x > 0$, and (iv) if $S_1(\nu_{h_2}(g) + h_1(g), h_2(g) + \varepsilon I_k)$ is absolutely continuous for all $g \in \mathcal{G}$, then $\int S_1(\nu_{h_2}(g) + h_1(g), h_2(g) + \varepsilon I_k)dQ(g)$ is absolutely continuous.

Assumption SQ(b) holds for $S_1$ because (i) if $v \geq 1$, the summand $\int \sum_{j=p+1}^{k}(\nu_{h_2,j}^2(g) /(h_{2,j,j}(g) + \varepsilon))dQ(g)$ has positive density on $[0, \infty)$, and (ii) if $v = 0$ and $h_1(g) \neq \infty^p$ on some $G \subset \mathcal{G}$ such that $Q(G) > 0$, each summand $\int [\nu_{h_2,j}(g) + h_{1,j}(g)]_-^2/(h_{2,j,j}(g) + \varepsilon)dQ(g)$ for which $h_{1,j}(g) < \infty$ on some $G \subset \mathcal{G}$ such that $Q(G) > 0$ has positive density on $[0, \infty)$ and so does the sum over $\sum_{j=1}^{p}$.

For $S_2$, if $v = 0$ and $h_1(g) = \infty^p$ a.s. $[Q]$, then $S_2(\nu_{h_2}(g) + h_1(g), h_2(g) + \varepsilon I_k) = 0$ a.s. $[Q]$, $J_{(h_1,h_2)}(x) = 1$ for all $x > 0$, Assumption SQ(a) holds, and Assumption SQ(b) does not impose any restriction. Otherwise, $v \geq 1$ or $h_1(g) < \infty^p$ on a subset $G \subset \mathcal{G}$ such that $Q(G) > 0$. In this case, the random variable $\int S_2(\nu_{h_2}(g) + h_1(g), h_2(g) + \varepsilon I_k)dQ(g)$ has support $[0, \infty)$ and is absolutely continuous. Hence, Assumptions SQ(a)-(b) hold. $\square$

The proof of Theorem B6 uses the following Lemma.

**Lemma D1.** *Suppose Assumptions* M *and* S1 *hold. Then, for all* $h \in \mathcal{H}$*, under any sequence* $\{(\theta_n, F_n) : n \geq 1\} \in Seq^b(h_1^*, h)$,

$$T_n(\theta_n) \to_d \int S(\nu_{h_2}(g) + h_1(g), h_2(g) + \varepsilon I_k) dQ(g) \sim J_{(h_1,h_2)} \text{ as } n \to \infty.$$

**Comment.** Condition (iv) of $Seq^b(h_1^*, h)$ is not needed for the result of Lemma D1 to hold.

**Proof of Theorem B6.** First, we prove part(a). Suppose $\{(\theta_n, F_n) : n \geq 1\} \in Seq^b$. Then, there exist $h \in \mathcal{H}$ and $h_1^* \in \mathcal{H}_1^*(h)$ such that $\{(\theta_n, F_n) : n \geq 1\} \in Seq^b(h_1^*, h)$. We need to show that under $\{(\theta_n, F_n) : n \geq 1\}$, $\limsup_{n\to\infty} P_{F_n}(T_n(\theta_n) \leq c_{n,b}(\theta_n, 1-\alpha)) \geq 1-\alpha$. The asymptotic distribution of $T_n(\theta_n)$ is given by Lemma D1. We now determine the probability limit of $c_{n,b}(\theta_n, 1-\alpha)$.

Let $J_{(h_1,h_2)}(x)$ for $x \in R$ denote the distribution function of $J_{(h_1,h_2)}$. By Lemma 5 in Andrews and Guggenberger (2010), if (i) $U_{n,b}(\theta_n, x) \to_p J_{(h_1^*,h_2)}(x)$ for all $x \in C(J_{(h_1^*,h_2)})$, where $C(J_{(h_1^*,h_2)})$ denotes the continuity points of $J_{(h_1^*,h_2)}$, and (ii) for all $\xi > 0$, $J_{(h_1^*,h_2)}(c_\infty + \xi) > 1 - \alpha$, where $c_\infty$ is the $1-\alpha$ quantile of $J_{(h_1^*,h_2)}$, then

$$c_{n,b}(\theta_n, 1-\alpha) \to_p c_\infty. \tag{15.42}$$

Condition (i) holds by the properties of U-statistics of degree $b$ and $T_{n,b,j}(\theta_n) \to_d J_{(h_1^*,h_2)}$ (see Thm. 2.1(i) in Politis and Romano (1994)). The latter holds by Lemma D1 because subsample $j$ is an i.i.d. sample of size $b$ from the population distribution.

By Assumption S1(c), $J_{(h_1,h_2)}(x) = 0 \ \forall x < 0$ for $h \in \mathcal{H}$. Thus, $c_\infty \geq 0$. If $v = 0$ and $h_1(g) = \infty^p$ a.s.[$Q$], then $J_{(h_1^*,h_2)}(0) = 1$, $c_\infty = 0$, $J_{(h_1^*,h_2)}(c_\infty + \xi) = 1 > 1-\alpha$. In all other cases, Assumption SQ(b) applies, $J_{(h_1^*,h_2)}(0) < 1$, and $J_{(h_1^*,h_2)}(c_\infty + \xi) > J_{(h_1^*,h_2)}(c_\infty) \geq 1-\alpha$. Thus, condition (ii) holds and (15.42) is established.

If $c_\infty > 0$, $c_\infty \in C(J_{(h_1,h_2)})$ by Assumption SQ(a). Thus,

$$\liminf_{n\to\infty} P_{F_n}(T_n(\theta_n) \leq c_{n,b}(\theta_n, 1-\alpha)) = J_{(h_1,h_2)}(c_\infty) \geq J_{(h_1^*,h_2)}(c_\infty) = 1-\alpha, \tag{15.43}$$

where the first equality holds by (15.42) and Lemma D1, the inequality holds by Assumption S1(b) and $h_1^* \leq h_1$, and the second equality holds by Assumption SQ(a) and the definition of $c_\infty$.

If $c_\infty = 0$, for some set $G \subset \mathcal{G}$ with $Q(G) = 1$, we have

$$
\begin{aligned}
& P_{F_n}(T_n(\theta_n) \leq c_{n,b}(\theta_n, 1-\alpha)) \\
\geq\ & P_{F_n}(T_n(\theta_n) \leq 0) \\
=\ & P_{F_n}\left(\frac{n^{1/2}\overline{m}_{n,j}(\theta_n, g)}{\overline{\sigma}_{n,j}(\theta_n, g)} \geq 0\ \forall j \leq p\ \&\ \frac{\overline{m}_{n,j}(\theta_n, g)}{\overline{\sigma}_{n,j}(\theta_n, g)} = 0\ \forall j = p+1, ..., k, \forall g \in G\right) \\
\to\ & P\left(\frac{\nu_{h,j}(g) + h_{1,j}(g)}{h_{2,j,j}(g) + \varepsilon} \geq 0\ \forall j \leq p\ \&\ \frac{\nu_{h,j}(g)}{h_{2,j,j}(g) + \varepsilon} = 0\ \forall j = p+1, ..., k, \forall g \in G\right) \\
=\ & P(S(\nu_h(g) + h_1(g), h_2(g) + \varepsilon I_k) = 0\ \forall g \in G) \\
=\ & J_{(h_1,h_2)}(0) \geq J_{(h_1^*,h_2)}(0) \geq 1 - \alpha, \qquad (15.44)
\end{aligned}
$$

where $\overline{\sigma}_{n,j}(\theta, g)$ and $h_{2,j,j}(g)$ denote the $j$th diagonal elements of $\overline{\Sigma}_n(\theta, g)$ and $h_2(g)$, respectively. In (15.44), the first inequality holds because $c_{n,b}(\theta_n, 1-\alpha)$ is the $1-\alpha$ sample quantile of the subsample test statistics and the test statistics are non-negative (by Assumption S1(a)), the first and second equalities hold by Assumption S2, the convergence holds by Lemma A1(a)-(b), the third equality holds by the definition of $J_{(h_1,h_2)}$, and the last inequality holds because 0 is the $1-\alpha$ quantile of $J_{(h_1^*,h_2)}$.

Next, we prove part (b). Let $(\theta_n^*, F_n^*) = (\theta, F)$ for $n \geq 1$, where $(\theta, F)$ is specified in Assumption C. Then, $\{(\theta_n^*, F_n^*) : n \geq 1\} \in Seq^b(h_1^*, h)$, where $h_1^* = h_{1,F}(\theta)$ and $h = (h_{1,F}(\theta), h_{2,F}(\theta))$. Thus,

$$
\liminf_{n\to\infty} P_{F_n^*}(T_n(\theta_n^*) \leq c_{n,b}(\theta_n^*, 1-\alpha)) = J_{(h_1,h_2)}(c_\infty) = J_{(h_1^*,h_2)}(c_\infty) = 1 - \alpha. \qquad (15.45)
$$

This and the result of Theorem B6(a) establish part (b).

Lastly, we prove part (c). Suppose Assumption Sub holds and $\{(\theta_{m_n}, F_{m_n}) : n \geq 1\}$ belongs to $Seq^b$ (where $Seq^b$ is defined with $m_n$ in place of $n$). Then,

$$
\begin{aligned}
AsyCS &= \lim_{n\to\infty} P_{F_{m_n}}(T_n(\theta_{m_n}) \leq c_{n,b}(\theta_{m_n}, 1-\alpha)) \\
&\geq \inf_{\{(\theta_n, F_n): n\geq 1\} \in Seq^b} \liminf_{n\to\infty} P_{F_n}(T_n(\theta_n) \leq c_{n,b}(\theta_n, 1-\alpha)) \\
&= 1 - \alpha \qquad (15.46)
\end{aligned}
$$

using Theorem B6(b). On the other hand,

$$\begin{aligned}
AsyCS &= \liminf_{n\to\infty} \inf_{(\theta,F)\in\mathcal{F}} P_F(T_n(\theta) \le c_{n,b}(\theta, 1-\alpha)) \\
&\le \inf_{\{(\theta_n,F_n):n\ge 1\}\in Seq^b} \liminf_{n\to\infty} P_{F_n}(T_n(\theta_n) \le c_{n,b}(\theta_n, 1-\alpha)) \\
&= 1-\alpha. \qquad (15.47)
\end{aligned}$$

Thus, we have $AsyCS = 1-\alpha$. $\square$

**Proof of Lemma D1.** By the same argument as used above to show (14.20), but with $\nu_{\widehat{h}_{2,n}(\theta_c)}(g)$ and $\varphi_n(\theta_c, g)$ replaced by $\nu_{n,F_n}(\theta_n, g)$ and $h_{1,n,F_n}(\theta_n, g)$, respectively, we have

$$T_n(\theta_n) \to_d T(h) = \int S(\nu_{h_2}(g) + h_1(g), h_2(g) + \varepsilon I_k) dQ(g), \qquad (15.48)$$

where $\nu_{n,F_n}(\theta_n, \cdot) \Rightarrow \nu_{h_2}(\cdot)$ by Lemma A1(a), $h_{1,n,F_n}(\theta_n, g) \to h_1(g)\ \forall g \in \mathcal{G}$ by Definition $Seq^b(h_1^*, h)$(ii), and $d(\widehat{h}_{2,n}(\theta_n), h_2) \to 0$ by Lemma A1(b) and (12.26). Note that the assumption that $\{(\theta_n, F_n) : n \ge 1\}$ satisfies Definition $Seq^b(h_1^*, h)$ and Assumption M implies that $\{(\theta_n, F_n) : n \ge 1\}$ satisfies Definition $SubSeq(h_2)$ and hence the conditions of Lemma A1 hold. $\square$

# 16 Supplemental Appendix E

This Appendix proves Lemma A1, which is stated in Supplemental Appendix A.

## 16.1 Preliminary Lemmas E1-E3

Before we prove Lemma A1, we review a few concepts from Pollard (1990) and state several lemmas that are used in the proof.

**Definition E1 (Pollard, 1990, Definition 3.3).** *The packing number $D(\xi, \rho, G)$ for a subset $G$ of a metric space $(\mathcal{G}, \rho)$ is defined as the largest $b$ for which there exist points $g^{(1)}, ..., g^{(b)}$ in $G$ such that $\rho(g^{(s)}, g^{(s')}) > \xi$ for all $s \neq s'$. The covering number $N(\xi, \rho, G)$ is defined to be the smallest number of closed balls with $\rho$-radius $\xi$ whose union covers $G$.*

It is easy to see that $N(\xi, \rho, G) \leq D(\xi, \rho, G) \leq N(\xi/2, \rho, G)$.

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be the underlying probability space equipped with probability distribution $\mathbf{P}$. Let $\{f_{n,i}(\omega, g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ be a triangular array of random processes. Let

$$\mathcal{F}_{n,\omega} = \{(f_{n,1}(\omega, g), ..., f_{n,n}(\omega, g))' : g \in \mathcal{G}\}. \tag{16.1}$$

Because $\mathcal{F}_{n,\omega} \subset R^n$, we use the Euclidean metric $||\cdot||$ on this space. For simplicity, we omit the metric argument in the packing number function, i.e., we write $D(\xi, G)$ in place of $D(\xi, ||\cdot||, G)$ when $G \subset \mathcal{F}_{n,\omega}$.

Let $\odot$ denote the element-by-element product. For example for $a, b \in R^n$, $a \odot b = (a_1 b_1, ..., a_n b_n)'$. Let **envelope functions** of a triangular array of processes $\{f_{n,i}(\omega, g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ be an array of functions $\{F_n(\omega) = (F_{n,1}(\omega), ..., F_{n,n}(\omega))' : n \geq 1\}$ such that $|f_{n,i}(\omega, g)| \leq F_{n,i}(\omega)$ $\forall i \leq n, n \geq 1, g \in \mathcal{G}, \omega \in \Omega$.

**Definition E2 (Pollard, 1990, Definition 7.9).** *A triangular array of processes* $\{f_{n,i}(\omega, g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ is said to be *manageable with respect to envelopes* $\{F_n(\omega) : n \geq 1\}$ *if there exists a deterministic real function $\lambda$ on* $(0, 1]$ *for which* (i) $\int_0^1 \sqrt{\log \lambda(\xi)} d\xi < \infty$ *and* (ii) $D(\xi ||\alpha \odot F_n(\omega)||, \alpha \odot \mathcal{F}_{n,\omega}) \leq \lambda(\xi)$ *for* $0 < \xi \leq 1$, *all* $\omega \in \Omega$, *all $n$-vectors $\alpha$ of nonnegative weights, and all* $n \geq 1$.

**Lemma E1.** *If a row-wise i.i.d. triangular array of random processes $\{\phi_{n,i}(\omega, g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ is manageable with respect to the envelopes $\{F_n(\omega) : n \geq 1\}$ and $c_n(\omega) = (c_{n,1}(\omega), ..., c_{n,n}(\omega))'$ is an $R^n$-valued function on the underlying probability* *space, then*

(a) $\{\phi_{n,i}(\omega,g)c_{n,i}(\omega) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ *is manageable with respect to the envelopes*

$$F_n(\omega) = (F_{n,1}(\omega)|c_{n,1}(\omega)|, ..., F_{n,n}(\omega)|c_{n,n}(\omega)|)' \text{ for } n \geq 1, \tag{16.2}$$

(b) $\{E\phi_{n,i}(\cdot,g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ *is manageable with respect to the envelopes* $\{EF_n : n \geq 1\}$ *provided* $EF_{n,1} < \infty$ *for all* $n \geq 1$, *and*

(c) *if another triangular array of random processes* $\{\phi^*_{n,i}(\omega,g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ *is manageable with respect to the envelopes* $\{F^*_n(\omega) : n \geq 1\}$, *then* $\{\phi^*_{n,i}(\omega,g)+\phi_{n,i}(\omega,g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ *is manageable with respect to the envelopes* $\{F_n(\omega)+F^*_n(\omega) : n \geq 1\}$.

**Lemma E2.** *If the triangular array of processes* $\{f_{n,i}(\omega,g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ *is manageable with respect to the envelopes* $\{F_n(\omega) = (F_{n,1}(\omega), ..., F_{n,n}(\omega))' : n \geq 1\}$, *and there exist* $0 < \eta < 1$ and $0 < B^* < \infty$ *such that* $n^{-1}\sum_{i \leq n} EF^{1+\eta}_{n,i} \leq B^*$ *for all* $n \geq 1$, *then*

$$\sup_{g \in \mathcal{G}} \left| n^{-1}\sum_{i=1}^{n}(f_{n,i}(\omega,g) - Ef_{n,i}(\cdot,g)) \right| \rightarrow_p 0. \tag{16.3}$$

Lemma E1(b)-(c) imply that if $\{f_{n,i}(\omega,g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ is manageable, then the triangular array of recentered processes $\{f_{n,i}(\omega,g) - Ef_{n,i}(\cdot,g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ also is manageable with respect to their corresponding envelopes. Lemma E2 is a uniform weak law of large numbers for triangular arrays of row-wise independent random processes. Lemma E2 is a complement to Thm. 8.2 in Pollard (1990) which is a uniform weak law of large numbers for independent sequences of random processes.

Lemma A1(a) is a functional central limit theorem result for multi-dimensional empirical processes. We prove it using a functional central limit theorem for real-valued empirical processes given in Pollard (1990, Thm. 10.3) and the Cramér-Wold device.

For $a \in R^k/\{0_k\}$, let

$$\begin{aligned} f_{n,i}(\omega,g) = {} & a'D^{-1/2}_{F_n}(\theta_n)n^{-1/2}[m(W_{n,i}(\omega),\theta_n,g) - E_{F_n}m(W_{n,i}(\cdot),\theta_n,g)], \\ & \text{for } \omega \in \Omega,\ g \in \mathcal{G}, \end{aligned} \tag{16.4}$$

where $W_{n,i}(\cdot) = W_i$, and the index $n$ in $W_{n,i}$ signifies the fact that the distribution of $W_i$ is changing with $n$. The random variable $f_{n,i}(\omega,g)$ depends on $a$, but for notational simplicity, $a$ does not appear explicitly in $f_{n,i}(\omega, g)$. By definition, we have

$$a'\nu_{n,F_n}(\theta_n, g) = \sum_{i=1}^{n} f_{n,i}(\omega, g). \tag{16.5}$$

Let

$$\rho_{n,a}(g, g^*) = \left(nE|f_{n,i}(\cdot, g) - f_{n,i}(\cdot, g^*)|^2\right)^{1/2} \text{ for } g, g^* \in \mathcal{G}. \tag{16.6}$$

We show in the proof of Lemma E3 below that under the assumptions, the sequence $\{\rho_{n,a}(g, g^*) : n \geq 1\}$ converges for each pair $g, g^* \in \mathcal{G}$. In consequence, the pointwise limit of $\rho_{n,a}(\cdot, \cdot)$ is an appropriate choice for the pseudo-metric on $\mathcal{G}$. Denote the limit by $\rho_a(\cdot, \cdot)$, i.e.,

$$\rho_a(g, g^*) = \lim_{n \to \infty} \rho_{n,a}(g, g^*). \tag{16.7}$$

**Lemma E3.** *For all $a \in R^k/\{0\}$ and any subsequence $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\} \in SubSeq(h_2)$*, for some $k \times k$-matrix-valued covariance kernel $h_2$ on $\mathcal{G} \times \mathcal{G}$,

(a) *$\mathcal{G}$ is totally bounded under the pseudo-metric $\rho_a$,*

(b) *the finite dimensional distributions of $a'\nu_{a_n,F_{a_n}}(\theta_{a_n}, g)$ have Gaussian limits with zero means and covariances given by $a'h_2(g, g^*)a$, $\forall g, g^* \in \mathcal{G}$, which uniquely determine a Gaussian distribution $\nu_a$ concentrated on the space of uniformly $\rho_a(\cdot, \cdot)$-continuous bounded functionals on $\mathcal{G}$, $U_{\rho_a}(\mathcal{G})$, and*

(c) *$a'\nu_{a_n,F_{a_n}}(\theta_{a_n}, \cdot)$ converges in distribution to $\nu_a$.*

The proofs of Lemmas E1-E3 are given below. The proof of Lemma E2 uses the maximal inequality in (7.10) of Pollard (1990). The proof of Lemma E3 uses the real-valued empirical process result of Thm. 10.6 in Pollard (1990).

## 16.2 Proof of Lemma A1(a)

Lemma A1 is stated in terms of subsequences $\{a_n\}$. For notational simplicity, we prove it for the sequence $\{n\}$. All of the arguments in this subsection and the next go through with $\{a_n\}$ in place of $\{n\}$.

The following three conditions are sufficient for weak convergence: (a) $(\mathcal{G}, \rho)$ is a totally bounded pseudo-metric space, (b) finite dimensional convergence holds: $\forall\{g^{(1)}, ..., g^{(L)}\} \subset \mathcal{G}$, $(\nu_{n,F_n}(\theta_n, g^{(1)})', ..., \nu_{n,F_n}(\theta_n, g^{(L)})')'$ converges in distribution, and

(c) $\{\nu_{n,F_n}(\theta_n,\cdot) : n \geq 1\}$ is stochastically equicontinuous. (For example, see Thm. 10.2 of Pollard (1990).)

First, we establish the total boundedness of the pseudo-metric space $(\mathcal{G},\rho)$, i.e., $N(\xi,\rho,\mathcal{G}) < \infty$ for all $\xi > 0$. This is done by constructing a finite collection of closed balls that covers $(\mathcal{G},\rho)$.

Consider $\xi > 0$. Let $B_\rho(g,\xi)$ denote a closed ball centered at $g$ with $\rho$-radius $\xi$. Let $\#G$ denote the number of elements in $G$ when $G$ is a finite set. (Throughout this proof $G$ denotes a subset of $\mathcal{G}$, not the envelope function that appears in Assumption M.) For $j = 1,...,k$, let $e_j$ be a $k$-dimensional vector with the $j$th coordinate equal to one and all other coordinates equal to zero. Then, $e_j \in R^k/\{0\}$ and by Lemma E3(a), the pseudo-metric spaces $(\mathcal{G},\rho_{e_j})$ are totally bounded. Consequently, for all $G \subset \mathcal{G}$, $(G,\rho_{e_j})$ is totally bounded. Our construction of the collection of closed balls is based on the following relationship between $\{\rho_{e_j} : j \leq k\}$ and $\rho$: $\forall g, g^* \in \mathcal{G}$,

$$\begin{aligned}\rho^2(g,g^*) &= tr\left(h_2(g,g) - h_2(g,g^*) - h_2(g^*,g) + h_2(g^*,g^*)\right)\\ &= \lim_{n\to\infty} E_{F_n}||D_{F_n}^{-1/2}(\theta_n)[\widetilde{m}(W_i,\theta_n,g) - \widetilde{m}(W_i,\theta_n,g^*)]||^2\\ &= \lim_{n\to\infty}\sum_{j=1}^k \rho^2_{n,e_j}(g,g^*) = \sum_{j=1}^k \rho^2_{e_j}(g,g^*), \end{aligned} \tag{16.8}$$

where the second equality holds by (16.7), which is proved in (16.40)-(16.41).

We start with $j = 1$. Because $(\mathcal{G},\rho_{e_1})$ is totally bounded, we can find a set $G_1 \subset \mathcal{G}$ such that

$$\#G_1 = N(\xi_k,\rho_{e_1},\mathcal{G}) \text{ and } \sup_{g\in\mathcal{G}}\min_{g^*\in G_1}\rho_{e_1}(g,g^*) \leq \xi_k, \tag{16.9}$$

where $\xi_k = \xi/(2\sqrt{k})$. For all $g \in G_1$, let $B^1_{\rho_{e_1}}(g,\xi_k) = B_{\rho_{e_1}}(g,\xi_k)\cap\mathcal{G}$. Then, $\bigcup_{g\in G_1} B^1_{\rho_{e_1}}(g,\xi_k)$ covers $\mathcal{G}$.

Because $B^1_{\rho_{e_1}}(g,\xi_k) \subset \mathcal{G}$, $(B^1_{\rho_{e_1}}(g,\xi_k),\rho_{e_2})$ is totally bounded. We are then able to choose a set $G_{2,g}$ such that

$$\#G_{2,g} = N(\xi_k,\rho_{e_2},B^1_{\rho_{e_1}}(g,\xi_k)) \text{ and } \sup_{g'\in B^1_{\rho_{e_1}}(g,\xi_k)}\min_{g^*\in G_{2,g}}\rho_{e_2}(g',g^*) \leq \xi_k. \tag{16.10}$$

Let $G_2 = \bigcup_{g\in G_1} G_{2,g}$. We have $\#G_2 = \sum_{g\in G_1}\#G_{2,g} < \infty$. For all $g \in G_1$ and $g' \in G_{2,g}$, let

$$B^2_{\rho_{e_2}}(g',\xi_k) = B_{\rho_{e_2}}(g',\xi_k) \cap B^1_{\rho_{e_1}}(g,\xi_k). \tag{16.11}$$

By construction, $\bigcup_{g' \in G_{2,g}} B^2_{\rho_{e_2}}(g', \xi_k)$ covers $B^1_{\rho_{e_1}}(g, \xi_k)$. Because $\bigcup_{g \in G_1} B^1_{\rho_{e_1}}(g, \xi_k)$ covers $\mathcal{G}$, $\bigcup_{g' \in G_2} B^2_{\rho_{e_2}}(g', \xi_k)$ covers $\mathcal{G}$.

Repeat the previous steps to obtain in turn $G_3$, $\{B^3_{\rho_{e_3}}(g, \xi_k) : g \in G_3\}$, ..., $G_k$, $\{B^k_{\rho_{e_k}}(g, \xi_k) : g \in G_k\}$. One can induce that (i) $\#G_k < \infty$, (ii) $\bigcup_{g' \in G_k} B^k_{\rho_{e_k}}(g', \xi_k)$ covers $\mathcal{G}$, and (iii) $\forall g \in \mathcal{G}$, there exists $(g^{(k)}, g^{(k-1)}, ..., g^{(1)}) \in G_k \times G_{k-1} \times ... \times G_1$ such that

$$g \in B^k_{\rho_{e_k}}(g^{(k)}, \xi_k) \subset B^{k-1}_{\rho_{e_{k-1}}}(g^{(k-1)}, \xi_k) \subset ... \subset B^1_{\rho_{e_1}}(g^{(1)}, \xi_k). \tag{16.12}$$

Thus,

$$\rho(g, g^{(k)}) = \left( \sum_{j=1}^{k} \rho^2_{e_j}(g, g^{(k)}) \right)^{1/2} \leq \left( \frac{\xi^2}{4k} + \frac{4\xi^2}{4k} + ... + \frac{4\xi^2}{4k} \right)^{1/2} < \xi. \tag{16.13}$$

Equation (16.13) implies that $\bigcup_{g \in G_k} B^k_{\rho}(g, \xi)$ covers $\mathcal{G}$, $G_k$ is the desired finite collection we set out to construct, $N(\xi, \rho, \mathcal{G}) \leq \#G_k < \infty$, and $(\mathcal{G}, \rho)$ is totally bounded.

Second, we show that finite dimensional convergence holds. By Lemma E3, the finite dimensional random vector $(a'\nu_{n,F_n}(\theta_n, g^{(1)}), ..., a'\nu_{n,F_n}(\theta_n, g^{(I)}))'$ converges in distribution:

$$\begin{pmatrix} a'\nu_{n,F_n}(\theta_n, g^{(1)}) \\ \vdots \\ a'\nu_{n,F_n}(\theta_n, g^{(L)}) \end{pmatrix} \to_d N \begin{pmatrix} 0, & a'h_2(g^{(1)}, g^{(1)})a & ... & a'h_2(g^{(1)}, g^{(L)})a \\ & \vdots & ... & \vdots \\ & a'h_2(g^{(L)}, g^{(1)})a & ... & a'h_2(g^{(L)}, g^{(L)})a \end{pmatrix} \tag{16.14}$$

for all $a \in R^k$. Thus, by the Cramér-Wold device, for all $g^{(1)}, g^{(2)}, ..., g^{(L)} \in \mathcal{G}$,

$$\begin{pmatrix} \nu_{n,F_n}(\theta_n, g^{(1)}) \\ \vdots \\ \nu_{n,F_n}(\theta_n, g^{(L)}) \end{pmatrix} \to_d N \left( 0, \begin{pmatrix} h_2(g^{(1)}, g^{(1)}) & ... & h_2(g^{(1)}, g^{(L)}) \\ \vdots & ... & \vdots \\ h_2(g^{(L)}, g^{(1)}) & ... & h_2(g^{(L)}, g^{(L)}) \end{pmatrix} \right). \tag{16.15}$$

Lastly, we show that $\{\nu_{n,F_n}(\theta_n, \cdot) : n \geq 1\}$ is stochastically equicontinuous with respect to $\rho$. By Lemma E3, $\{e'_j\nu_{n,F_n}(\theta_n, \cdot) : n \geq 1\}$ is stochastically equicontinuous with respect to $\rho_{e_j}$ for all $j \leq k$. (Weak convergence implies stochastic equicontinuity.) Because $\rho(g, g^*) \geq \rho_{e_j}(g, g^*)$ for all $g, g^* \in \mathcal{G}$, $\{e'_j\nu_{n,F_n}(\theta_n, \cdot) : n \geq 1\}$ is stochastically equicontinuous with respect to $\rho$ for all $j \leq k$. Note that $e'_j\nu_{n,F_n}(\theta_n, \cdot)$ is the $j$th coordinate of $\nu_{n,F_n}(\theta_n, \cdot)$. Therefore, $\{\nu_{n,F_n}(\theta_n, \cdot) : n \geq 1\}$ is stochastically equicontinuous with respect to $\rho$. $\square$

## 16.3 Proof of Lemma A1(b)

It suffices to show that each element of $D_F^{-1/2}(\theta)\widehat{\Sigma}_n(\theta,g,g^*)D_F^{-1/2}(\theta)$ converges in probability uniformly over $g,g^*\in\mathcal{G}$. Suppose $1\le j,j'\le k$. The $(j,j')$th element of $D_{F_n}^{-1/2}(\theta_n)\widehat{\Sigma}_n(\theta_n,g,g^*)D_{F_n}^{-1/2}(\theta_n)$ can be decomposed into two parts:

$$\begin{aligned}
&n^{-1}\sum_{i=1}^n \sigma_{F_n,j}^{-1}(\theta_n)m_j(W_i,\theta_n)m_{j'}(W_i,\theta_n)\sigma_{F_n,j'}^{-1}(\theta_n)g_j\left(X_i\right)g_{j'}^*\left(X_i\right)\\
&-\sigma_{F_n,j}^{-1}(\theta_n)\overline{m}_{n,j}(\theta_n,g)\overline{m}_{n,j'}(\theta_n,g^*)\sigma_{F_n,j'}^{-1}(\theta_n)\\
\equiv\ & n^{-1}\sum_{i=1}^n f_{n,i,j,j'}^{mm}(\omega,g,g^*)-n^{-1}\sum_{i=1}^n f_{n,i,j}^m(\omega,g)\left(n^{-1}\sum_{i=1}^n f_{n,i,j'}^m(\omega,g^*)\right), \quad (16.16)
\end{aligned}$$

where

$$\begin{aligned}
f_{n,i,j}^m(\omega,g) &= \sigma_{F_n,j}^{-1}(\theta_n)m_j(W_i,\theta_n)g_j\left(X_i\right), \text{ and}\\
f_{n,i,j,j'}^{mm}(\omega,g,g^*) &= f_{n,i,j}^m(\omega,g)f_{n,i,j'}^m(\omega,g^*). \quad (16.17)
\end{aligned}$$

Note that $\{f_{n,i,j,j'}^{mm}(\omega,g,g^*):g,g^*\in\mathcal{G},i\le n,n\ge 1\}$ and $\{f_{n,i,j}^m(\omega,g):g\in\mathcal{G},i\le n,n\ge 1\}$ are triangular arrays of row-wise i.i.d. random processes. We show the uniform convergence of their sample means using Lemma E2.

We first study $f_{n,i,j}^m(\omega,g)$. Let

$$\mathcal{F}_{n,\omega,j}^m=\{(f_{n,1,j}^m(\omega,g),...,f_{n,n,j}^m(\omega,g))':g\in\mathcal{G}\}. \quad (16.18)$$

By Assumption M(c) and Lemma E1, $\{f_{n,i,j}^m(\omega,g):i\le n,\ g\in\mathcal{G}\}$ are manageable with respect to the envelopes

$$\begin{aligned}
F_{n,\cdot,j}^m(\omega) &= \left(F_{n,1,j}^m(\omega),...,F_{n,n,j}^m(\omega)\right)', \text{ where}\\
F_{n,i,j}^m(\omega) &= G(X_i)\sigma_{F_n,j}^{-1}(\theta_n)|m_j(W_i,\theta_n)|. \quad (16.19)
\end{aligned}$$

In consequence, there exist functions $\lambda_j:(0,1]\to[0,\infty)$ for $j\le k$ such that

$$D\left(\xi|\alpha\odot F_{n,\cdot,j}^m|,\alpha\odot\mathcal{F}_{n,\omega,j}^m\right)\le\lambda_j(\xi) \quad (16.20)$$

for all $\alpha \in [0,\infty)^n$, $\omega \in \Omega$, and $n \geq 1$ and $\sqrt{\log \lambda_j(\xi)}$ is integrable over $(0,1]$.

Because the data are i.i.d., we have for all $0 < \eta \leq 1$ and all $n$,

$$
\begin{aligned}
& n^{-1} \sum_{i=1}^{n} E(F^m_{n,i,j})^{1+\eta} = E(F^m_{n,1,j})^{1+\eta} \\
\leq\; & \left(E_{F_n} G^{\delta_1}(X_i)\right)^{(1+\eta)/\delta_1} \left( E_{F_n} \left| \frac{m_j\left(W_1, \theta_n\right)}{\sigma_{F_{n,h},j}(\theta_n)} \right|^{\delta_2} \right)^{(1+\eta)/\delta_2} < \infty, \qquad (16.21)
\end{aligned}
$$

where $\delta_2 = (1+\eta)\delta_1/(\delta_1 - 1 - \eta)$. The first inequality above holds by Hölder's inequality and the second holds by Assumption M(b), $\delta_2 \leq 2 + 4/(\delta_1 - 1 - \eta) \leq 2 + 4/(4\delta^{-1} + 1 - \eta) \leq 2 + \delta$, and condition (vi) of (2.3). Therefore, by Lemma E2,

$$
\sup_{g \in \mathcal{G}} \left| n^{-1} \sum_{i=1}^{n} f^m_{n,i,j}(\omega, g) - E f^m_{n,1,j}(\cdot, g) \right| \to_p 0. \qquad (16.22)
$$

Now we study $f^{mm}_{n,i,j,j'}(\omega, g, g^*)$. For all $n \geq 1$ and $\omega \in \Omega$, let

$$
\mathcal{F}^{mm}_{n,\omega,j,j'} = \{(f^{mm}_{n,1,j,j'}(\omega, g, g^*), ..., f^{mm}_{n,n,j,j'}(\omega, g, g^*))' : g, g^* \in \mathcal{G}\}. \qquad (16.23)
$$

Then, $\mathcal{F}^{mm}_{n,\omega,j,j'} = \mathcal{F}^m_{n,\omega,j} \odot \mathcal{F}^m_{n,\omega,j}$. Let $F^{mm}_{n,\cdot,j,j'}(\omega) = F^m_{n,\cdot,j}(\omega) \odot F^m_{n,\cdot,j'}(\omega)$. We have: for all $\alpha \in [0,\infty)^n$, $\omega \in \Omega$, and $n \geq 1$,

$$
\begin{aligned}
& D\left(\xi | \alpha \odot F^{mm}_{n,\cdot,j,j'}(\omega)|, \alpha \odot \mathcal{F}^{mm}_{n,\omega,j,j'}\right) \\
=\; & D\left(\xi | \alpha \odot F^{mm}_{n,\cdot,j,j'}(\omega)|, \alpha \odot \mathcal{F}^m_{n,\omega,j} \odot \mathcal{F}^m_{n,\omega,j}\right) \\
\leq\; & D\left(\frac{\xi}{4} | \alpha \odot F^m_{n,\cdot,j'}(\omega) \odot F^m_{n,\cdot,j}(\omega)|, \alpha \odot F^m_{n,\cdot,j'}(\omega) \odot \mathcal{F}^m_{n,\omega,j}\right) \\
& \cdot D\left(\frac{\xi}{4} | \alpha \odot F^m_{n,\cdot,j}(\omega) \odot F^m_{n,\cdot,j'}(\omega)|, \alpha \odot F^m_{n,\cdot,j}(\omega) \odot \mathcal{F}^m_{n,\omega,j'}\right) \\
\leq\; & \lambda_j(\xi/4)\lambda_{j'}(\xi/4), \qquad (16.24)
\end{aligned}
$$

where the first inequality holds by equation (5.2) in Pollard (1990) and the second inequality holds by (16.20). We have

$$\int_0^1 \sqrt{\log(\lambda_j(\xi/4)\lambda_{j'}(\xi/4))}d\xi = \int_0^1 \sqrt{\log\lambda_j(\xi/4) + \log\lambda_{j'}(\xi/4)}d\xi$$
$$\leq 4\int_0^{1/4} \left(\sqrt{\log\lambda_j(\xi)} + \sqrt{\log\lambda_{j'}(\xi)}\right) d\xi < \infty, \tag{16.25}$$

where the first inequality holds by $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$. Therefore, $\{f^{mm}_{n,i,j,j'}(\omega, g, g^*) : g, g^* \in \mathcal{G}, i \leq n, n \geq 1\}$ are manageable with respect to the envelopes $\{F^{mm}_{n,\cdot,j,j'}(\omega) : n \geq 1\}$.

Let $\eta$ be a small positive number. We have

$$\begin{aligned}
&n^{-1}\sum_{i\leq n} E(F^{mm}_{n,j,j'}(\cdot))^{1+\eta} = E(F^{mm}_{n,j,j'}(\cdot))^{1+\eta} \\
&\leq \left[E_{F_n} G^{\delta_3}(X_1)\right]^{2(1+\eta)/\delta_3} \left[E_{F_n}\left|\frac{m_j\left(W_1,\theta_n\right)}{\sigma_{F_n,j}(\theta_n)}\right|^{2+\delta}\right]^{(1+\eta)/(2+\delta)} \\
&\quad \cdot \left[E_{F_n}\left|\frac{m_{j'}\left(W_1,\theta_n\right)}{\sigma_{F_n,j'}(\theta_n)}\right|^{2+\delta}\right]^{(1+\eta)/(2+\delta)} \\
&< \infty,
\end{aligned} \tag{16.26}$$

where $\delta_3 = 2(1+\eta)(2+\delta)/(\delta - 2\eta)$, the first inequality holds by Hölder's inequality, and the second holds for sufficiently small $\eta > 0$ by Assumption M(b) and condition (vi) of (2.3).

With the manageability of $\{f^{mm}_{n,i,j,j'}(\omega, g, g^*) : g, g^* \in \mathcal{G}, i \leq n, n \geq 1\}$ and (16.26), Lemma E2 gives

$$\sup_{g,g^*\in\mathcal{G}} \left| n^{-1}\sum_{i=1}^n f^{mm}_{n,i,j,j'}\left(\omega, g, g^*\right) - Ef^{mm}_{n,1,j,j'}\left(\cdot, g, g^*\right)\right| \to_p 0. \tag{16.27}$$

By (16.16), (16.22), (16.27), as well as $|Ef^{mm}_{n,1,j}(\cdot, g)| \leq E(F^m_{n,1,j})^{1+\eta} < \infty$, we conclude that the difference between the $(j,j')$th element of $D^{-1/2}_{F_n}(\theta_n)\widehat{\Sigma}_n(\theta_n, g, g^*)D^{-1/2}_{F_n}(\theta_n)$ and $Ef^{mm}_{n,1,j,j'}(\cdot, g, g^*) - Ef^m_{n,1,j}(\cdot, g)Ef^m_{n,1,j'}(\cdot, g^*)$ converges to zero uniformly over $(g, g^*) \in \mathcal{G}^2$.

By definition,

$$
\begin{aligned}
& Ef^{mm}_{n,1,j,j'}(\cdot,g,g^*) - Ef^{m}_{n,1,j}(\cdot,g)Ef^{m}_{n,1,j'}(\cdot,g^*) \\
= \; & E_{F_n}[\sigma^{-1}_{F_n,j}(\theta_n)\sigma^{-1}_{F_n,j'}(\theta_n)m_j\left(W_1,\theta_n\right)g_j\left(X_1\right)m_{j'}(W_1,\theta_n)g^*_{j'}(X_1)] \\
& -E_{F_n}\left[\sigma^{-1}_{F_n,j}(\theta_n)m_j(W_1,\theta_n)g_j(X_1)\right]E_{F_n}[\sigma^{-1}_{F_n,j'}(\theta_n)m_{j'}(W_1,\theta_n)g^*_{j'}(X_1)] \\
= \; & \sigma^{-1}_{F_n,j}(\theta_n)\sigma^{-1}_{F_n,j'}(\theta_n)\left[\Sigma_{F_n}(\theta_n,g,g^*)\right]_{j,j'} \\
\to \; & [h_2(g,g^*)]_{j,j'},
\end{aligned}
\tag{16.28}
$$

where the convergence holds uniformly over $(g,g^*) \in \mathcal{G}^2$ by conditions (i) and (iv) in Definition $SubSeq(h_2)$. This completes the proof of Lemma A1(b). $\square$

## 16.4 Proof of Lemma E1

Part (a) is proved by a similar, but simpler, argument to that given in (16.24)-(16.25).

Next, we prove part (b). Because $EF_{n,i} < \infty$ and the processes $\{\phi_{n,i}(\omega,g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ are row-wise i.i.d., $E\mathcal{F}_n \equiv \{E\phi_{n,i}(\cdot,g)\cdot 1_n : g \in \mathcal{G}\}$ is a subset of a one dimensional affine subspace of $R^n$ with diameter no greater than $2EF_{n,i}$. Thus, $\alpha \odot E\mathcal{F}_n$ is a subset of a one dimensional affine subspace of $R^n$ with diameter no greater than $2||\alpha||EF_{n,i}$. By Lem. 4.1 in Pollard (1990), we have: for all $n \geq 1$,

$$
D\left(\xi||\alpha \odot EF_n||, \alpha \odot E\mathcal{F}_n\right) \leq 6||\alpha||EF_{n,i}/(\xi||\alpha \odot EF_n||) = 6/\xi. \tag{16.29}
$$

Because $\int_0^1 \sqrt{\log(6/\xi)}d\xi = 3\sqrt{\pi} < \infty$, part (b) holds.

Finally, we prove part (c). Let $\lambda^*_\phi(\xi) : (0,1] \to R^+$ be the square-root-log integrable function such that

$$
D\left(\xi\left\|\alpha \odot F^*_n(\omega)\right\|, \alpha \odot \mathcal{F}^*_{n,\omega}\right) \leq \lambda^*_\phi\left(\xi\right) \text{ for } 0 < \xi \leq 1, \tag{16.30}
$$

for all $\alpha \in [0,\infty)^n$, $\omega \in \Omega$, and $n \geq 1$. Let

$$
\begin{aligned}
\mathcal{F}^*_{n,\omega} &= \{\phi^*_n(\omega,g) : g \in \mathcal{G}\}, \\
\mathcal{F}^{sum}_{n,\omega} &= \{\phi_n(\omega,g) + \phi^*_n(\omega,g) : g \in \mathcal{G}\}, \text{ and} \\
\mathcal{F}^+_{n,\omega} &= \mathcal{F}^*_{n,\omega} \oplus \mathcal{F}_{n,\omega} \equiv \{a+b \in R^n : a \in \mathcal{F}_{n,\omega}, b \in \mathcal{F}^*_{n,\omega}\},
\end{aligned}
\tag{16.31}
$$

where $\phi_n(\omega, g) = (\phi_{n,1}(\omega, g), ..., \phi_{n,n}(\omega, g))'$. Let

$$F_n^{sum}(\omega) = F_n(\omega) + F_n^*(\omega). \tag{16.32}$$

Then, for $0 < \xi \leq 1$ and $\alpha \in [0, \infty)^n$,

$$\begin{aligned}
&D\left(\xi||\alpha \odot F_n^{sum}(\omega)||, \alpha \odot \mathcal{F}_{n,\omega}^{sum}\right) \\
&\leq D\left(\xi||\alpha \odot F_n^{sum}(\omega)||, \alpha \odot \mathcal{F}_{n,\omega}^{+}\right) \\
&\leq D\left(\xi(||\alpha \odot F_n(\omega)|| + ||\alpha \odot F_n^*(\omega)||)/\sqrt{2}, \alpha \odot \mathcal{F}_{n,\omega}^{+}\right) \\
&\leq D(\xi||\alpha \odot F_n(\omega)||/(2\sqrt{2}), \alpha \odot \mathcal{F}_{n,\omega}) \\
&\quad \cdot D(\xi||\alpha \odot F_n^*(\omega)||/(2\sqrt{2}), \alpha \odot \mathcal{F}_{n,\omega}^*) \\
&\leq \lambda_\phi(\xi/(2\sqrt{2}))\lambda_\phi^*(\xi/(2\sqrt{2})),
\end{aligned} \tag{16.33}$$

where $\lambda_\phi(\xi)$ denotes the packing number bounding function given in Definition E2 for the processes $\{\phi_n(\omega, g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$, the first inequality holds because $\mathcal{F}_{n,\omega}^{sum} \subset \mathcal{F}_{n,\omega}^{+}$, the second inequality holds because $D(x, G)$ is decreasing in $x$ and $||a+b|| \geq (||a||+||b||)/\sqrt{2}$ for $a, b \in [0, \infty)^n$, the third inequality holds by a stability result for packing numbers (see Pollard (1990, p. 22)), and the last inequality holds by the manageability of $\{\phi_n(\omega, g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ and (16.30).

The function $\lambda_\phi(\xi/(2\sqrt{2}))\lambda_\phi^*(\xi/(2\sqrt{2}))$ is square-root-log integrable by (16.25), which completes the proof of part (c). $\square$

## 16.5 Proof of Lemma E2

We prove convergence in probability by showing convergence in $L^1$. We have

$$\begin{aligned}
&E\sup_{g\in\mathcal{G}}\left|n^{-1}\sum_{i=1}^n [f_{n,i}(\cdot, g) - Ef_{n,i}(\cdot, g)]\right| \leq n^{-1}KE\left(\sum_{i=1}^n F_{n,i}^2\right)^{1/2} \\
&\leq n^{-1}KE\left(\sum_{i=1}^n F_{n,i}^{1+\eta}\right)^{1/(1+\eta)} \leq n^{-1}K\left(E\sum_{i=1}^n F_{n,i}^{1+\eta}\right)^{1/(1+\eta)} \\
&\leq n^{-\eta/(1+\eta)}K\left(B^*\right)^{1/(1+\eta)} \to 0 \text{ as } n \to \infty,
\end{aligned} \tag{16.34}$$

where the first inequality holds for some constant $K < \infty$ by manageability and the maximal inequality (7.10) in Pollard (1990), the second inequality holds using $0 < \eta < 1$

by applying the inequality $\sum_{i=1}^n x_i^s \leq (\sum_{i=1}^n x_i)^s$, which holds for $s \geq 1$ and $x_i \geq 0$ for $i = 1, ..., n$, with $x_i = F_{n,i}^{1+\eta}$ and $s = 2/(1+\eta) > 0$, the third inequality holds by the concavity of the function $f(x) = x^{1/(1+\eta)}$ when $\eta > 0$, and the last inequality holds because $n^{-1} \sum_{i=1}^n EF_{n,1}^{1+\eta} \leq B^*$ for all $n \geq 1$. $\square$

## 16.6 Proof of Lemma E3

For notational simplicity, we prove Lemma E3 for the sequence $\{n\}$, rather than the subsequence $\{a_n\}$. All of the arguments in this subsection go through with $\{a_n\}$ in place of $\{n\}$.

The conclusions of Lemma E3 are implied by the result of Thm. 10.6 of Pollard (1990), which relies on the following five conditions:

(i) the $\{f_{ni}(\omega, g) : g \in \mathcal{G}\}$ defined in (16.4) are manageable with respect to some envelope $F_{a,n}(\omega) = (F_{a,n,1}(\omega), ..., F_{a,n,n}(\omega))'$,

(ii) $\lim_{n\to\infty} Ea'\nu_{n,F_n}(\theta_n, g)\nu_{n,F_n}(\theta_n, g^*)'a = a'h_2(g, g^*)a$ for all $g$, $g^* \in \mathcal{G}$,

(iii) $\limsup_{n\to\infty} \sum_{i=1}^n EF_{a,n,i}^2 < \infty$,

(iv) $\sum_{i=1}^n EF_{a,n,i}^2\{F_{a,n,i} > \xi\} \to 0$ as $n \to \infty$ for each $\xi > 0$, and

(v) the limit $\rho_a(\cdot, \cdot)$ is well defined by (16.7), and for all deterministic sequences $\{g_{(n)}\}$ and $\{g_{(n)}^*\}$, if $\rho_a(g_{(n)}, g_{(n)}^*) \to 0$, then $\rho_{n,a}(g_{(n)}, g_{(n)}^*) \to 0$ as $n \to \infty$.

Now we verify the five conditions.

(i) By (16.4), we have

$$\begin{aligned} f_{n,i}(\omega, g) = \sum_{j=1}^k a_j \sigma_{F_n,j}^{-1}(\theta_n) n^{-1/2} [&m_j(W_{n,i}(\omega), \theta_n) g_j(X_{n,i}(\omega)) \\ &-E_{F_n} m_j(W_i, \theta_n) g_j(X_i)], \end{aligned} \tag{16.35}$$

where $a_j$ denotes the $j$th element of $a$. By Assumption M(c), $\{g_j(X_{n,i}(\omega)) : i \leq n\}$ are manageable with respect to envelopes $G(X_{n,i}(\omega))$. Therefore, by Lemma E1(a)-(c), $\{f_{n,i}(\omega, g) : i \leq n\}$ is manageable with respect to envelopes $F_{a,n} = (F_{a,n,1}, ..., F_{a,n,n})'$ defined by

$$\begin{aligned} F_{a,n,i}(\omega) = n^{-1/2} \sum_{j=1}^k a_j \sigma_{F_n,j}^{-1}(\theta_n) [&|m_j(W_{n,i}(\omega), \theta_n)| G(X_{ni}(\omega)) \\ &+E_{F_n}|m_j(W_i, \theta_n)| G(X_i)]. \end{aligned} \tag{16.36}$$

(ii) By (16.5), we have

$$
\begin{aligned}
& Ea'\nu_{n,F_n}(\theta_n, g)\nu'_{n,F_n}(\theta_n, g^*)a \\
= {} & E\left(\sum_{i=1}^{n} f_{n,i}(\cdot, g)\right)\left(\sum_{i=1}^{n} f_{n,i}\left(\cdot, g^*\right)\right)' = nEf_{n,1}(\cdot, g)f_{n,1}(\cdot, g^*)' \\
= {} & n^{-1}a'D_{F_n}^{-1/2}(\theta_n) \cdot Cov_{F_n}\left(m(W_1, \theta_n, g), m(W_1, \theta_n, g^*)\right) \cdot D_{F_n}^{-1/2}(\theta_n)a \\
= {} & n^{-1}a'D_{F_n}^{-1/2}(\theta_n)\Sigma_{F_n}(\theta_n, g, g^*)D_{F_n}^{-1/2}(\theta_n)a, \qquad (16.37)
\end{aligned}
$$

where the second equality holds because the data are i.i.d., the third inequality holds by (16.4). Condition (i) in Definition $SubSeq(h_2)$ completes the verification of condition (ii) above.

(iii) Next, we verify $\limsup_{n\to\infty} \sum_{i=1}^{n} EF_{a,n,i}^2 < \infty$. By the linear structure of $F_{a,n,i}$, it suffices to show that

$$
\begin{aligned}
& \limsup_{n\to\infty} E_{F_n}\sigma_{F_n,j}^{-2}(\theta_n)|m_j(W_i, \theta_n)|^2 G^2(X_i) < \infty \text{ and} \\
& \limsup_{n\to\infty} E_{F_n}\sigma_{F_n,j}^{-1}(\theta_n)|m_j(W_i, \theta_n)|G(X_i) < \infty. \qquad (16.38)
\end{aligned}
$$

The latter is implied by the former and the former holds by the same argument as in (16.21) with $\eta = 1$.

(iv) For $B$ as in condition (vi) of (2.3), $\xi > 0$, and $\eta > 0$ sufficiently small,

$$
\begin{aligned}
& \sum_{i=1}^{n} EF_{a,n,i}^2\{F_{a,n,i} > \xi\} = nEF_{a,n,i}^2\{F_{a,n,i} > \xi\} \le nEF_{a,n,i}^{2+\eta}/\xi^{\eta} \\
\le {} & \frac{2\left(2k\right)^{2+\eta}}{n^{\eta/2}\xi^{\eta}} \sum_{j=1}^{k} |a_j|^{2+\eta} E_{F_n} G^{2+\eta}(X_i)\sigma_{F_n,j}^{-2-\eta}(\theta_n)|m_j\left(W_i, \theta_n\right)|^{2+\eta} \\
\le {} & \frac{2\left(2k\right)^{2+\eta}}{n^{\eta/2}\xi^{\eta}} \sum_{j=1}^{k} |a_j|^{2+\eta} \left[E_{F_n} G^{\delta_4}(X_1)\right]^{(2+\eta)/\delta_4} B^{(2+\eta)/(2+\delta)} \\
\le {} & \frac{2\left(2k\right)^{2+\eta} B^{(2+\eta)/(2+\delta)} C^{(2+\eta)/\delta_1}}{n^{\eta/2}\xi^{\eta}} \sum_{j=1}^{k} |a_j|^{2+\eta} \to 0, \qquad (16.39)
\end{aligned}
$$

where the first equality holds because the data are identically distributed, the second inequality holds by Jensen's inequality using the convexity of $\psi\left(x\right) = x^{2+\eta}$, i.e., $((2k)^{-1}\sum_{j=1}^{k}(|X_j| + E|X_j|))^{2+\eta} \le (2k)^{-1}\sum_{j=1}^{k}(|X_j|^{2+\eta} + (E|X_j|)^{2+\eta})$ and $(E|X_j|)^{2+\eta} \le$

$E|X_j|^{2+\eta}$, the third inequality holds with $\delta_4 = (2+\eta)(2+\delta)/(\delta-\eta)$ by the same arguments as in (16.26), and the fourth inequality holds by Assumption M(b) and $\delta_4 \leq \delta_1$ for sufficiently small $\eta$.

(v) First we show that the limit $\rho_a(\cdot,\cdot)$ is well defined by (16.7). For any $g, g^* \in \mathcal{G}$,

$$\begin{aligned} \rho_{n,a}^2(g,g^*) &= nE(f_{n,i}(\cdot,g) - f_{n,i}(\cdot,g^*))^2 \\ &= a' D_{F_n}^{-1/2}(\theta_n) Var_{F_n}\left(m(W_i,\theta_n,g) - m(W_i,\theta_n,g^*)\right) D_{F_n}^{-1/2}(\theta_n) a \\ &\to a'h_2(g,g)a + a'h_2(g^*,g^*)a - a'h_2(g,g^*)a - a'h_2(g^*,g)a, \end{aligned} \tag{16.40}$$

where the convergence hold uniformly over $\mathcal{G}^2$ by condition (i) in Definition $SubSeq(h_2)$. Thus, $\rho_a(g,g^*) = \lim_{n\to\infty} \rho_{n,a}(g,g^*)$ is well defined, and

$$\lim_{n\to\infty} \sup_{g,g^*\in\mathcal{G}} \left|\rho_{n,a}(g,g^*) - \rho_a(g,g^*)\right| = 0. \tag{16.41}$$

Lastly, we show the second property of condition (v). Let $\xi > 0$ be arbitrary. Suppose $\rho_a(g_{(n)}, g^*_{(n)}) \to 0$. Then, there exists an $N_0 < \infty$ such that for $n \geq N_0$,

$$\rho_a(g_{(n)}, g^*_{(n)}) \leq \xi/2. \tag{16.42}$$

By (16.41), we have

$$\lim_{m\to\infty} \sup_{n\geq 1} \left|\rho_{m,a}(g_{(n)}, g^*_{(n)}) - \rho_a(g_{(n)}, g^*_{(n)})\right| = 0. \tag{16.43}$$

Thus, there exists an $N_1 < \infty$ such that for all $m \geq N_1$,

$$\sup_{n\geq 1} \left|\rho_{m,a}(g_{(n)}, g^*_{(n)}) - \rho_a(g_{(n)}, g^*_{(n)}\right| \leq \xi/2. \tag{16.44}$$

Take $N = \max\{N_0, N_1\}$, then we have for $n \geq N$,

$$\rho_{n,a}(g_{(n)}, g^*_{(n)}) \leq \xi. \tag{16.45}$$

Thus, $\rho_a(g_{(n)}, g^*_{(n)}) \to 0$ implies $\rho_{n,a}(g_{(n)}, g^*_{(n)}) \to 0$. $\square$

# 17 Supplemental Appendix F

This Appendix provides additional material concerning the Monte Carlo simulations in the quantile selection and entry game models in Sections 17.1 and 17.4. In addition, it provides all of the Monte Carlo simulation results for the mean selection and interval-outcome regression models in Sections 17.2 and 17.3.

## 17.1 Quantile Selection Model

The first subsection of this section provides additional simulation results to those given in the paper. The second subsection provides figures for the conditional moment functions evaluated at the $\theta$ values at which the FCP's are computed in Table IV of the paper. The third subsection describes the computation of the Chernozhukov, Lee, and Rosen (2008) (CLR) and Lee, Song, and Whang (2011) (LSW) CI's.

### 17.1.1 Additional Simulation Results

Table S-I provides comparisons of the coverage probability (CP) and false coverage probability (FCP) performance of the CvM and KS test statistics and PA and GMS critical values in the quantile selection model with *peaked bound function*. These comparisons are analogous to those reported in Table I of the paper for the flat and kinked bound functions. The results for the peaked bound are similar to those for the flat and kinked bound functions except that there is little difference between the FCP's for the CvM and KS versions of the test statistics.

Table S-I. Quantile Selection Model: Base Case Test Statistic Comparisons for Peaked Bound Function*

| | | | | | |
|---|---|---|---|---|---|
| (a) Coverage Probabilities | | | | | |
| | Statistic: | CvM/Sum | CvM/Max | KS/Sum | KS/Max |
| DGP | Crit Val | | | | |
| Peaked Bd | PA/Asy | 1.000 | 1.000 | .997 | .997 |
| | GMS/Asy | .997 | .997 | .991 | .990 |
| (b) False Coverage Probabilities (coverage-probability corrected) | | | | | |
| Peaked Bd | PA/Asy | .70 | .68 | .48 | .47 |
| | GMS/Asy | .43 | .41 | .39 | .38 |

* These results are for the lower endpoint of the identified interval. They are based on (5000, 5001) CP (and FCP) and critical value repetitions, respectively.

Table S-II provides coverage probability (CP) and false coverage probability (FCP) results for the upper endpoint of the identified interval in the quantile selection model.[54] (Table I of AS provides analogous results for the lower endpoint.) Table S-II provides a comparison of CS's based on the CvM/Sum, CvM/QLR, CvM/Max, KS/Sum, KS/QLR, and KS/Max statistics, coupled with the PA/Asy and GMS/Asy critical values. The relative attributes of the different CS's are quite similar to those reported in Table I of AS for the lower endpoint. None of the CS's under-cover. So, the relative attributes of the CS's are determined by their FCP's. The CvM-based CS's have lower FCP's than the KS-based CS's. The CS's that use the GMS/Asy critical values have lower FCP's than those based on the PA/Asy critical values. The FCP's do not depend on whether the Sum, QLR, or Max version of the statistic is employed. Hence, the best CS of those considered is the CvM/Max/GMS/Asy CS, or this CS with Max replaced by Sum or QLR.

Table S-II. Quantile Selection Model, Upper Endpoint: Base Case Test Statistic Comparisons

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | (a) Coverage Probabilities | | | | | |
| | Statistic: | CvM/Sum | CvM/QLR | CvM/Max | KS/Sum | KS/QLR | KS/Max |
| DGP | Crit Val | | | | | | |
| Flat Bound | PA/Asy | .994 | .994 | .993 | .984 | .984 | .982 |
| | GMS/Asy | .971 | .971 | .970 | .974 | .974 | .972 |
| Kinked Bound | PA/Asy | .996 | .996 | .996 | .989 | .989 | .988 |
| | GMS/Asy | .974 | .974 | .972 | .976 | .976 | .975 |
| | | (b) False Coverage Probabilities (coverage probability corrected) | | | | | |
| Flat Bound | PA/Asy | .73 | .72 | .71 | .70 | .70 | .69 |
| | GMS/Asy | .42 | .42 | .42 | .55 | .55 | .55 |
| Kinked Bound | PA/Asy | .73 | .73 | .72 | .74 | .74 | .73 |
| | GMS/Asy | .41 | .41 | .41 | .52 | .52 | .52 |

[54] For the upper endpoint with the flat bound and the upper endpoint with the kinked bound, the FCP's are computed at the points $\underline{\theta}(1) + 0.40 \times sqrt(250/n)$ and $\underline{\theta}(1) + 0.75 \times sqrt(250/n)$, respectively. These points are chosen to yield similar values for the FCP's across the different cases considered.

Table S-III reports CP and FCP results for variations on the base case for the lower endpoint with the kinked bound DGP. (Table III of AS reports analogous results for the lower endpoint with the flat bound.) The results are similar to those in Table III of AS. There is relatively little sensitivity to the sample size, the number of cubes $g$, and the choice of $\varepsilon$. There is relatively little sensitivity of the CP's to the choice of $(\kappa_n, B_n)$, but some sensitivity of the FCP's with the base case choice being superior to values of $(\kappa_n, B_n)$ that are twice or half as large. The CS with $\alpha = .5$ is half-median unbiased and avoids the well-known problem of inward-bias. But, it is farther from being median-unbiased than in the flat bound case.

Table S-III. Quantile Selection Model, Kinked Bound, and Lower Endpoint: Variations on the Base Case

| | | (a) Coverage Probabilities | | (b) False CP's (CP-corrected) | |
|---|---|---|---|---|---|
| | Statistic: | CvM/Max | KS/Max | CvM/Max | KS/Max |
| Case | Crit Val: | GMS/Asy | GMS/Asy | GMS/Asy | GMS/Asy |
| Base Case ($n = 250, r_1 = 7, \varepsilon = 5/100$) | | .983 | .984 | .34 | .52 |
| $n = 100$ | | .981 | .985 | .34 | .55 |
| $n = 500$ | | .984 | .984 | .39 | .54 |
| $n = 1000$ | | .984 | .980 | .41 | .54 |
| $r_1 = 5$ | | .981 | .981 | .34 | .49 |
| $r_1 = 9$ | | .983 | .986 | .35 | .55 |
| $r_1 = 11$ | | .984 | .987 | .36 | .60 |
| $(\kappa_n, B_n) = 1/2(\kappa_{n,bc}, B_{n,bc})$ | | .984 | .997 | .39 | .51 |
| $(\kappa_n, B_n) = 2(\kappa_{n,bc}, B_{n,bc})$ | | .990 | .991 | .38 | .59 |
| $\varepsilon = 1/100$ | | .981 | .981 | .34 | .56 |
| $\alpha = .5$ | | .721 | .710 | .03 | .06 |
| $\alpha = .5$ & $n = 500$ | | .741 | .734 | .04 | .08 |

### 17.1.2 Conditional Moment Function Figures

Figure S-1 shows the conditional moment functions $\beta(x, \theta)$ (defined in (10.6)), as functions of $x$, evaluated at the $\theta$ values 1.531, 1.181, and 1.151 at which the FCP's are computed in Table IV of the paper in the flat, kinked, and peaked cases, respectively.

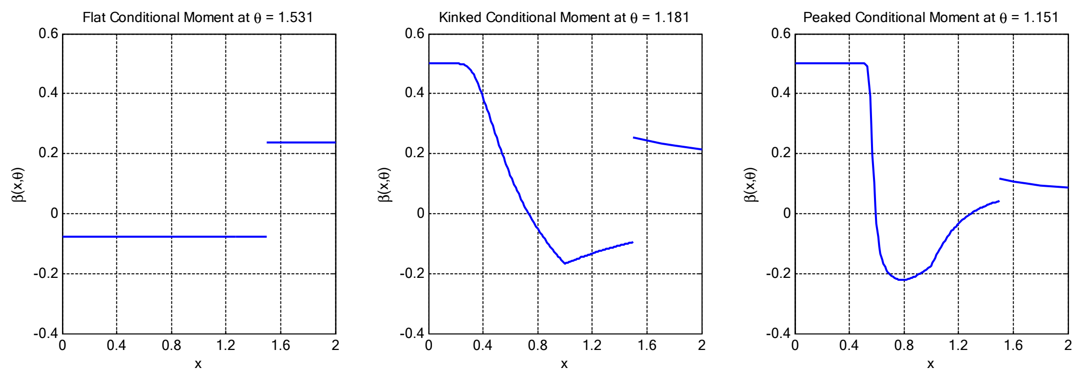


Figure S-1. Conditional Moment Functions for the Quantile Selection Model Evaluated at $\theta$ Values Below the Lower Endpoint of the Identified Set

### 17.1.3 Description of the CLR-Series, CLR-Local Linear, and LSW Confidence Intervals

Here we describe the computation of the CLR and LSW CI's reported in Table IV for the quantile selection model and Table S-V (given below) for the mean selection model. In the quantile selection model, the parameter $\theta$ is not separable from its bound functions. Thus, we handle the model following the method in Example 4 of CLR. We define an auxiliary parameter $\beta$:

$$\beta(\theta) = \min_{x \in R} \beta(x, \theta), \text{ where} \tag{17.1}$$

$$\beta(x, \theta) = \begin{cases} E\left(1(Y_i \leq \theta, T_i = t) + 1(T_i \neq t) - \tau | X_i = x\right) & \text{if } x < x_0 \\ E\left(\tau - 1(Y_i \leq \theta, T_i = t) | X_i = x\right) & \text{if } x \geq x_0. \end{cases} \tag{17.2}$$

We obtain a CLR bound estimator $\widehat{\beta}_\alpha(\theta)$ for a null $\theta$ value and let the nominal $1 - \alpha$ confidence set for $\theta$ be $CS_n^{CLR}(\alpha) = \{\theta : \widehat{\beta}_\alpha(\theta) \geq 0\}$. In the mean selection model, the parameter $\theta$ is separable from its bound function, so computation is as described in CLR.

We follow the procedure described on pp. 28-29 and 50-51 of CLR to compute $\widehat{\beta}_\alpha(\theta)$ with the following alterations: (1) for the standard error of the spline coefficients (the choice of which is not described in CLR), we use the Eicker-White formula, (2) for the set of numbers of spline functions considered in the cross-validation procedure, we increase the set to $\{5, 6, ..., 13\}$, and (3) to compute the many minima and maxima involved, we use a grid-search combined with Newton-Raphson method. Specifically, regarding the latter, we take 101 evenly spaced grid-points between $[0, 2]$ (the support of $x$), compute the objective functions at the 101 points, and choose the point that gives the highest value as the starting point for the Newton-Raphson routine. Because the objective functions have multiple sharp peaks, we believe that the combined procedure gives more precise optima than doing the grid search or the Newton-Raphson alone. CLR does not describe the procedure they use to obtain the minima and maxima. As in CLR, we use cross-validation to determine the number of series/bandwidth parameter.

To obtain the LSW confidence set, for each $\theta$, we use LSW's test for the null hypothesis: $H_0 : -\beta(x, \theta) \leq 0 \ \forall x \in \mathcal{X}$, and let the confidence set be all the $\theta$ values such that the test does not reject. We use the $L_1$-version of their test. We follow the descriptions on p. 9 of LSW and adopt the same tuning parameters (weight, kernel, bandwidth, etc.) as in their Monte Carlo simulation. We use 5000 random draws to simulate the mean and covariance of the Gaussian vectors appearing in their test statistic, and use the Gaussian quadrature method to carry out the numerical integration.

## 17.2 Mean Selection Model

In this section, we consider the same mean selection model that is considered in CLR. We compare the CP's and FCP's of the CI's based on the CvM and KS statistics and the PA and GMS critical values.[55] We also compare the CvM/Max/GMS/Asy CI (abbreviated by AS below) with several other CI's in the literature, viz., the CLR-series, CLR-local linear, and LSW CI's.[56]

The model is essentially the same as the quantile selection model described in the paper except that the parameter of interest $\theta$ is the conditional mean $E(y_i(1)|X_i = x_0)$ for some $x_0$, rather than the conditional quantile. In addition, the QMIV assumption

[55] These comparisons are similar to those given in Table I of the paper for the quantile selection model.

[56] These comparisons are similar to those given in Table IV of the paper for the quantile selection model.

is replaced with the monotone instrumental variable (MIV) assumption of Manski and Pepper (2000): for all $(x_1, x_2) \in \mathcal{X}^2$ such that $x_1 \leq x_2$,

$$E(y_i(1)|X_i = x_1) \leq E(y_i(1)|X_i = x_2). \tag{17.3}$$

The MIV assumption is not informative unless $y_i(t)$ has bounded support. Let the support of $y_i(1)$ be $[Y_l, Y_u]$. The MIV assumption leads to the following moment inequalities:

$$\begin{aligned} &E\left(1(X_i \leq x_0)[\theta - Y_i 1\left(T_i = 1\right) - Y_l 1(T_i \neq 1)]|X_i\right) \geq 0 \text{ a.s. and} \\ &E\left(1(X_i \geq x_0)[Y_i 1\left(T_i = 1\right) + Y_u 1(T_i \neq 1) - \theta]|X_i\right) \geq 0 \text{ a.s.} \end{aligned} \tag{17.4}$$

We consider the same data generating processes (DGP's) as in Section 4 of CLR. That is, $y_i(1) = \mu(X_i) + \sigma(X_i) u_i$ and $[Y_l, Y_u] = [-1.96, 1.96]$, where $X_i \sim Unif[-2, 2]$ and $u_i \sim 1.96 \wedge ((-1.96) \vee N(0,1))$, $T_i = 1\{L(X_i) + \varepsilon_i \geq 0\}$, where $\varepsilon_i \sim N(0,1)$ and $\varepsilon_i, u_i,$ and $X_i$ are independent of each other, and $Y_i = y_i(T_i)$. Two specifications of $(\mu(x), \sigma(x), L(x))$ are considered, which yield flat and kinked bound functions for the conditional mean $\theta$. For the flat bound DGP, $\mu(x) = 0 = L(x)$ and $\sigma(x) = |x|$. For the kinked bound DGP, $\mu(x) = 2(x \wedge 1)$, $L(x) = x \wedge 1$, and $\sigma(x) = |x|$. The DGP is the same as in (10.4) of the paper for the quantile selection model except for the distributions of $X_i$ and $u_i$. The parameter of interest is the conditional mean of $y_i(1)$ at $x_0 = 1.5$. That is, $\theta = E(y_i(1)|X_i = 1.5)$.

We consider sample size $n = 250$ (which is also the base case sample size for the quantile selection model in the paper). All results concern the lower end of the identified interval for $\theta$, which equals $-.98$ and $1.372$ in the flat and kinked bound cases, respectively.[57] All results are based on (5000, 5001) coverage probability and critical value repetitions, respectively. The FCP's are CP-corrected, as described in Section 10 of the paper.[58]

[57] The DGP is the same for FCP's as for CP's, just the value $\theta$ that is to be covered is different. For the lower endpoint of the identified set, FCP's are computed for $\theta$ equal to $\underline{\theta}(1) - c$, where $c = .155$ and $.68$ in the flat and kinked bound cases, respectively. These points are chosen to yield similar values for the FCP's across the two cases.

[58] That is, a positive constant is added to the critical value such that the CP for the given case being considered is .95 whenever the CP for the given case (without correction) is less than .95.

Table S-IV. Mean Selection Model: Base Case Test Statistic and Critical Value Comparisons

| (a) Coverage Probabilities (95%) | | | | | |
|---|---|---|---|---|---|
| | Statistic: | CvM/Sum | CvM/Max | KS/Sum | KS/Max |
| DGP | Crit Val | | | | |
| Flat Bd | PA/Asy | .976 | .972 | .974 | .970 |
| | GMS/Asy | .951 | .950 | .959 | .958 |
| Kinked Bd | PA/Asy | 1.000 | 1.000 | .997 | .997 |
| | GMS/Asy | .972 | .970 | .946 | .942 |
| (b) False Coverage Probabilities (coverage-probability corrected) | | | | | |
| Flat Bd | PA/Asy | .49 | .46 | .70 | .68 |
| | GMS/Asy | .38 | .37 | .63 | .63 |
| Kinked Bd | PA/Asy | .88 | .86 | .61 | .59 |
| | GMS/Asy | .39 | .38 | .33 | .33 |

Tables S-IV and S-V report the simulation results for the mean selection model. Table S-IV provides CP and FCP comparisons of the CI's based on the test statistics CvM/Sum, CvM/Max, KS/Sum, and KS/Max and the PA and GMS critical values. All results are for the "asymptotic" versions of the tests (whose critical values are determined by simulating the asymptotic distributions), not the bootstrap versions. The CP probability results are quite similar to those for the quantile selection model. The same is true for the FCP results for the flat bound function. For the kinked bound function, the main difference is that the CvM form of the test statistic does not out-perform the KS version, which it does in the quantile selection model. In particular, Table S-IV shows that the CvM/Max statistic combined with the GMS critical value performs very well. It has CP equal to .950 in the flat bound case and .970 in the kinked bound case. It has the lowest FCP in the flat bound case and close to the lowest FCP in the kinked bound case.

Table S-V compares the AS CI with the CLR-series, CLR-local linear, and LSW CI's. The AS and LSW CI's have good CP properties, viz., CP's greater than or equal to .95. On the other hand, the two CLR CI's have poor CP properties. They under-cover substantially. The AS CI has clearly the best FCP's for the flat bound case. For the kinked bound case, the CLR-local linear CI has best FCP's followed by the CLR-series and AS CI's. The LSW CI has poor FCP's. In sum, the AS CI has the best combined CP and FCP properties by a substantial margin in the mean selection model with $n = 250$.

We note that the results in Table S-V for the AS CI are quite similar to the results in Table IV for the quantile selection model. The same is true for the LSW CI except that its FCP's are worse in the mean selection model. In the kinked bound case, the CLR CI's perform noticeably worse in the mean selection model with $n = 250$ (compared to the quantile selection model) in terms of CP's and better in terms of FCP's.

Table S-V. Mean Selection Model: Comparisons of Andrews and Shi (2008) Confidence Intervals with Those Proposed in Chernozhukov, Lee, and Rozen (2008) and Lee, Song, and Whang (2011)

| | CP (95%) | | FCP (corrected) | | CP (50%) | |
|---|---|---|---|---|---|---|
| CS | Flat | Kinked | Flat | Kinked | Flat | Kinked |
| $n = 250$ | | | | | | |
| CvM/Max/GMS/Asy | .950 | .970 | .37 | .38 | .48 | .68 |
| CLR-series | .912 | .883 | .78 | .36 | .47 | .56 |
| CLR-local linear | .849 | .910 | .84 | .25 | .37 | .64 |
| LeeSongWhang | .977 | 1.000 | .64 | 1.00 | .76 | 1.00 |

## 17.3 Interval-Outcome Regression Model

### 17.3.1 Description of Model

Here we report simulation results for an interval-outcome regression model. This model has been considered by Manski and Tamer (2002, Sec. 4.5). It is a regression model where the outcome variable $Y_i^*$ is partially observed:

$$Y_i^* = \theta_1 + X_i\theta_2 + U_i, \text{ where } E(U_i|X_i) = 0 \text{ a.s., for } i = 1, ..., n. \tag{17.5}$$

One observes $X_i$ and an interval $[Y_{L,i}, Y_{U,i}]$ that contains $Y_i^*$: $Y_{L,i} = \lfloor Y_i \rfloor$ and $Y_{U,i} = \lfloor Y_i \rfloor + 1$, where $\lfloor x \rfloor$ denotes the integer part of $x$. Thus, $Y_i^* \in [Y_{L,i}, Y_{U,i}]$.

It is straightforward to see that the following conditional moment inequalities hold in this model:

$$E(\theta_1 + X_i\theta_2 - Y_{L,i}|X_i) \geq 0 \text{ a.s. and}$$
$$E(Y_{U,i} - \theta_1 - X_i\theta_2|X_i) \geq 0 \text{ a.s.} \tag{17.6}$$

In the simulation experiment, we take the true parameters to be $(\theta_1, \theta_2) = (1, 1)$ (without loss of generality), $X_i \sim U[0, 1]$, and $U_i \sim N(0, 1)$. We consider a base case sample size of $n = 250$, as well as $n = 100$, 500, and 1000.

The parameter $\theta = (\theta_1, \theta_2)$ is not identified. Figure S-1 shows the identified set. It is a parallelogram in $(\theta_1, \theta_2)$ space enclosed by thick solid lines with vertices at $(.5, 1), (.5, 2), (1.5, 0)$, and $(1.5, 1)$. The point $(1, 1)$ is the true parameter. The thin solid lines are the lower bounds defined by the first moment inequality and the dashed lines are the upper bounds defined by the second moment inequality.

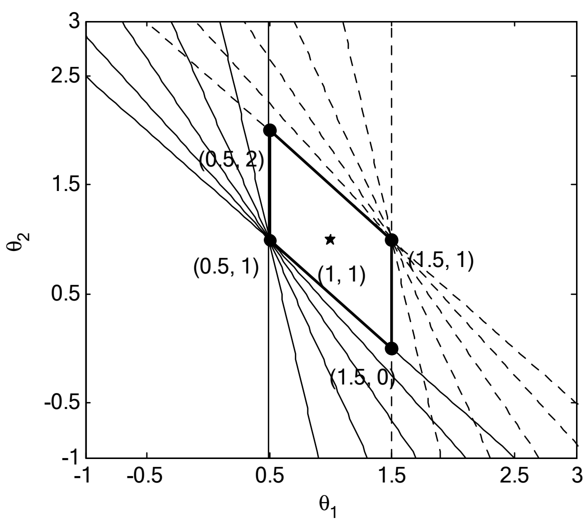


Figure S-2. The Identified Set of the Interval Outcome Model

By symmetry, CP's of CS's are the same for the points $(.5, 1)$ and $(1.5, 1)$. Also, they are the same for $(.5, 2)$ and $(1.5, 0)$. We focus on CP's at the corner point $(.5, 1)$, which is in the identified set, and at points close to $(.5, 1)$ but outside the identified set.[59]

[59] Specifically, the $\theta$ values outside the identified set are given by $\theta_1 = 0.5 - 0.075 \times (500/n)^{1/2}$ and $\theta_2 = 1.0 - 0.050 \times (500/n)^{1/2}$. These $\theta$ values are selected so that the FCP's of the CS's take values in an interesting range for all values of $n$ considered.

The corner point $(.5, 1)$ is of interest because it is a point in the identified set where CP's of CS's typically are strictly less than one. Due to the features of the model, the CP's of CS's typically equal one (or essentially equal one) at interior points, non-corner boundary points, and the corner points $(.5, 2)$ and $(1.5, 0)$.

#### 17.3.2 g Functions

The $g$ functions employed by the test statistics are indicator functions of hypercubes in $[0, 1]$. It is not assumed that the researcher knows that $X_i \sim U[0, 1]$ and so the regressor $X_i$ is transformed via the method described in Section 9 to lie in $(0, 1)$.[60] The hypercubes have side-edge lengths $(2r)^{-1}$ for $r = r_0, ..., r_1$, where $r_0 = 1$ and the base case value of $r_1$ is 7. The base case number of hypercubes is 56. We also report results for $r_1 = 5$, 9, and 11, which yield 30, 90, and 132 hypercubes, respectively. With $n = 250$ and $r_1 = 7$, the expected number of observations per cube is 125, $62.5, ..., 20.8$, or 17.9 depending on the cube. With $n = 250$ and $r_1 = 11$, the expected number also can equal 12.5 or 11.4. With $n = 100$ and $r_1 = 7$, the expected number is 50, $25, ..., 8.3$, or 7.3.

#### 17.3.3 Simulation Results

Tables S-VI, S-VII, and S-VIII provide results for the interval-outcome regression model that are analogous to the results in Tables I-III for the quantile selection model. In spite of the differences in the models—the former is linear and parametric with a bivariate parameter, while the latter is nonparametric with a scalar parameter—the results are similar.

Table S-VI shows that the CvM/Max statistic combined with the GMS/Asy critical value has CP's that are very close to the nominal level .95. Its FCP's are noticeably lower than those for CS's that use the KS form or PA-based critical values. The CvM/Sum-GMS/Asy and CvM/QLR-GMS/Asy CS's perform equally well as the Max version. Table S-VII shows that the results for the Asy and Bt versions of the critical values are quite similar for the CvM/Max-GMS CS, which is the best CS. The Sub critical value yields substantial under-coverage for the KS/Max statistic. The Sub critical values are dominated by the GMS critical values in terms of FCP's.

Table S-VIII shows that the CS's do not exhibit much sensitivity to the sample size or the number of cubes employed. It also shows that at the non-corner boundary point

[60] This method takes the transformed regressor to be $\Phi((X_i - \overline{X}_n)/\sigma_{X,n})$, where $\overline{X}_n$ and $\sigma_{X,n}$ are the sample mean and standard deviations of $X_i$ and $\Phi(\cdot)$ is the standard normal distribution function.

$\theta = (1.0, 0.5)$ and the corner point $\theta = (1.5, 0)$, all CP's are (essentially) equal to one.[61] Lastly, Table S-VIII shows that the lower endpoint estimator based on the CvM/Max-GMS/Asy CS with $\alpha = .5$ is close to being median-unbiased, as in the quantile selection model. It is less than the lower bound with probability is .472 and exceeds it with probability .528 when $n = 250$.

We conclude that the preferred CS for this model is of the CvM form, combined with the Max, Sum, or QLR function, and uses a GMS critical value, either Asy or Bt.

Table S-VI. Interval-Outcome Regression Model: Base Case Test Statistic Comparisons

| Critical Value | Statistic: | CvM/Sum | CvM/QLR | CvM/Max | KS/Sum | KS/QLR | KS/Max |
|---|---|---|---|---|---|---|---|
| | | (a) Coverage Probabilities | | | | | |
| PA/Asy | | .990 | .993 | .990 | .989 | .990 | .989 |
| GMS/Asy | | .950 | .950 | .950 | .963 | .963 | .963 |
| | | (b) False Coverage Probabilities (coverage probability corrected) | | | | | |
| PA/Asy | | .62 | .66 | .61 | .78 | .80 | .78 |
| GMS/Asy | | .37 | .37 | .37 | .61 | .61 | .61 |

[61] This is due to the fact that the CP's at these points are linked to their CP's at the corner point $\theta = (0.5, 1.0)$ given the linear structure of the model. If the CP is reduced at the two former points (by reducing the critical value), the CP at the latter point is very much reduced and the CS does not have the desired size.

Table S-VII. Interval-Outcome Regression Model: Base Case Critical Value Comparisons

(a) Coverage Probabilities

| Statistic | Critical Value: | PA/Asy | PA/Bt | GMS/Asy | GMS/Bt | Sub |
|---|---|---|---|---|---|---|
| CvM/Max | | .990 | .995 | .950 | .941 | .963 |
| KS/Max | | .989 | .999 | .963 | .953 | .890 |

(b) False Coverage Probabilities (coverage probability corrected)

| | | PA/Asy | PA/Bt | GMS/Asy | GMS/Bt | Sub |
|---|---|---|---|---|---|---|
| CvM/Max | | .61 | .69 | .37 | .38 | .45 |
| KS/Max | | .78 | .96 | .61 | .54 | .66 |

Table S-VIII. Interval-Outcome Regression Model: Variations on the Base Case

| | | (a) Coverage Probabilities | | (b) False Cov Probs (CPcor) | |
|---|---|---|---|---|---|
| | Statistic: | CvM/Max | KS/Max | CvM/Max | KS/Max |
| Case | Crit Val: | GMS/Asy | GMS/Asy | GMS/Asy | GMS/Asy |
| Base Case ($n = 250, r_1 = 7, \varepsilon = 5/100$) | | .950 | .963 | .37 | .61 |
| $n = 100$ | | .949 | .970 | .39 | .66 |
| $n = 500$ | | .950 | .956 | .37 | .60 |
| $n = 1000$ | | .954 | .955 | .37 | .60 |
| $r_1 = 5$ (30 cubes) | | .949 | .961 | .37 | .59 |
| $r_1 = 9$ (90 cubes) | | .951 | .965 | .37 | .63 |
| $r_1 = 11$ (132 cubes) | | .950 | .968 | .38 | .64 |
| $(\kappa_n, B_n) = 1/2(\kappa_{n,bc}, B_{n,bc})$ | | .944 | .961 | .40 | .62 |
| $(\kappa_n, B_n) = 2(\kappa_{n,bc}, B_{n,bc})$ | | .958 | .973 | .39 | .65 |
| $\varepsilon = 1/100$ | | .946 | .966 | .39 | .69 |
| $(\theta_1, \theta_2) = (1.0, 0.5)$ | | .999 | .996 | .91 | .92 |
| $(\theta_1, \theta_2) = (1.5, 0.0)$ | | 1.000 | .996 | .99 | .97 |
| $\alpha = .5$ | | .472 | .481 | .03 | .08 |
| $\alpha = .5$ & $n = 500$ | | .478 | .500 | .03 | .07 |

## 17.4 Entry Game Model

### 17.4.1 Probit Log Likelihood Function

In the entry game model, the probit log likelihood function for $\tau = (\tau_1, \tau_2)$ given $\theta = (\theta_1, \theta_2)$ is

$$\begin{aligned}
&\sum_{i=1}^{n} 1(Y_i = (0,0)) \ln(\Phi(-X_{i,1}'\tau_1)\Phi(-X_{i,2}'\tau_2)) \\
&+ \sum_{i=1}^{n} 1(Y_i = (1,1)) \ln(\Phi(X_{i,1}'\tau_1 - \theta_1)\Phi(X_{i,2}'\tau_2 - \theta_2)) \\
&+ \sum_{i=1}^{n} 1(Y_i = (1,0) \text{ or } Y_i = (0,1)) \ln(g_i(\tau,\theta)), \text{ where} \qquad (17.7) \\
&\quad g_i(\tau,\theta) = 1 - \Phi(-X_{i,1}'\tau_1)\Phi(-X_{i,2}'\tau_2) - \Phi(X_{i,1}'\tau_1 - \theta_1)\Phi(X_{i,2}'\tau_2 - \theta_2)
\end{aligned}$$

over $\tau \in R^8$ for fixed $\theta$. The estimator $\widehat{\tau}_n(\theta)$ maximizes this function over $\tau \in R^8$ given $\theta$.

The gradient of the probit log likelihood for $\tau$ given $\theta$ is

$$\begin{aligned}
&-\sum_{i=1}^{n} 1(Y_i = (0,0)) \begin{pmatrix} \psi(-X_{i,1}'\tau_1)X_{i,1} \\ \psi(-X_{i,2}'\tau_2)X_{i,2} \end{pmatrix} \\
&+\sum_{i=1}^{n} 1(Y_i = (1,1)) \begin{pmatrix} \psi(X_{i,1}'\tau_1 - \theta_1)X_{i,1} \\ \psi(X_{i,2}'\tau_2 - \theta_2)X_{i,2} \end{pmatrix} \\
&+\sum_{i=1}^{n} 1(Y_i = (1,0) \text{ or } Y_i = (0,1)) \frac{1}{g_i(\tau,\theta)} \qquad (17.8) \\
&\times \begin{pmatrix} \phi(-X_{i,1}'\tau_1)\Phi(-X_{i,2}'\tau_2)X_{i,1} - \phi(X_{i,1}'\tau_1 - \theta_1)\Phi(X_{i,2}'\tau_2 - \theta_2)X_{i,1} \\ \Phi(-X_{i,1}'\tau_1)\phi(-X_{i,2}'\tau_2)X_{i,2} - \Phi(X_{i,1}'\tau_1 - \theta_1)\phi(X_{i,2}'\tau_2 - \theta_2)X_{i,2} \end{pmatrix},
\end{aligned}$$

where $\psi(x) = \phi(x)/\Phi(x)$.

### 17.4.2 Identification

Here we briefly discuss identification of the entry game model. Tamer (2003, Thm. 1) provides identification results that cover the model considered in Section 10.3 because $X_{i,1}$ and $X_{i,2}$ both contain continuous regressors whose support is $R$.

We point out here that this support condition is probably much stronger than is needed for identification in many contexts. For example, suppose the unobservables $U_{i,1}$ and $U_{i,2}$ are independent and standard normal, as in Section 10.3. Suppose the regressor vectors are $X_{i,1} = (1, Z_i)'$ and $X_{i,2} = 1$ and their coefficient vectors are $\tau_1 = (\tau_{11}, \tau_{12})'$ and $\tau_2$, respectively. Then, $\tau_1$ and $\tau_2$ are identified provided $Z_i$ has a density with respect to Lebesgue measure on some non-degenerate interval and $\tau_{12} \neq 0$. Thus, in this case, no large support condition is needed.

To prove this result, note that $P(Y_i = (0,0)|X_{i,1}) = \Phi(-X'_{i,1}\tau_1)\Phi(-\tau_2)$. Thus, for identification at $(\tau_1, \tau_2)$, it suffices to show that

$$P(\Phi(-X'_{i,1}\tau_1)\Phi(-\tau_2) = \Phi(-X'_{i,1}\lambda_1)\Phi(-\lambda_2)) = 1 \tag{17.9}$$

only if $\lambda_1 = \tau_1$ and $\lambda_2 = \tau_2$.

Suppose $\lambda_2 = \tau_2$. Then, (17.9) holds iff $P(X'_{i,1}\tau_1 = X'_{i,1}\lambda_1) = 1$. The left-hand side equals $P(\tau_{11} - \lambda_{11} + Z_i(\tau_{12} - \lambda_{12}) = 0)$. Given the condition on $Z_i$, the latter equals one only if $\lambda_1 = \tau_1$. Hence, when $\lambda_2 = \tau_2$, $(\lambda_1, \lambda_2)$ is observational equivalent to $(\tau_1, \tau_2)$ only if $(\lambda_1, \lambda_2) = (\tau_1, \tau_2)$.

Next, suppose $\lambda_2 \neq \tau_2$. Let $c = \Phi(-\lambda_2)/\Phi(-\tau_2)$ $(\neq 1)$. Then, (17.9) holds iff $P(\Phi(-\tau_{11} - Z_i\tau_{12}) = \Phi(-\lambda_{11} - Z_i\lambda_{12})c) = 1$. The latter implies that for all $z$ in an open interval, say $I$, $\Phi(-\tau_{11} - z\tau_{12}) = \Phi(-\lambda_{11} - z\lambda_{12})c$. Taking the derivative with respect to $z$ for $z \in I$, one obtains $\phi(-\tau_{11} - z\tau_{12}) = \phi(-\lambda_{11} - z\lambda_{12})c\lambda_{12}/\tau_{12}$. Taking logs yields a quadratic equation in $z$ for $z \in I$:

$$\begin{aligned} &(\tau_{11} + z\tau_{12})^2 = (\lambda_{11} + z\lambda_{12})^2 + c_1 \text{ or} \\ &(\tau_{12}^2 - \lambda_{12}^2)z^2 + 2(\tau_{11}\tau_{12} - \lambda_{11}\lambda_{12})z + \tau_{11}^2 - \lambda_{11}^2 - c_1 = 0, \end{aligned} \tag{17.10}$$

where $c_1 = \log(c\lambda_{12}/\tau_{12})$ and $c_1$ is well-defined because $\tau_{12} \neq 0$. A quadratic equation cannot hold for all $z \in I$ unless each coefficient of the equation is zero because a non-degenerate quadratic equation has at most two solutions. Suppose $\tau_{12}^2 - \lambda_{12}^2 = 0$. Then, $\tau_{11}\tau_{12} - \lambda_{11}\lambda_{12} = 0$ requires $\tau_{11} = \pm\lambda_{11}$, which implies that $\tau_{11}^2 - \lambda_{11}^2 = 0$. In consequence, $\tau_{11}^2 - \lambda_{11}^2 - c_1 = -c_1 \neq 0$ and the quadratic equation is not degenerate. (Note that $c_1 \neq 0$ because $c_1 = 0$ iff $c\lambda_{12}/\tau_{12} = 1$ iff $\lambda_{12} = c\tau_{12}$, and the latter condition violates $\tau_{12}^2 - \lambda_{12}^2 = 0$.) In conclusion, if $\lambda_2 \neq \tau_2$, (17.9) cannot hold for any $\lambda_1$ and $\tau_1$. This completes the proof of identification.

Note that it is not clear that even continuity of $Z_i$ in a nondegenerate interval is necessary for identification of $\tau$. If $Z_i$ is discrete with $s \geq 3$ support points, then observational equivalence requires $s$ nonlinear equations in two unknowns to hold. These equations depend on the joint distribution $F(\cdot,\cdot)$ of $(U_{i,1}, U_{i,2})$. This suggests (but does not prove) that for most joint distribution functions $F(\cdot,\cdot)$ of $(U_{i,1}, U_{i,2})$ identification holds under quite weak conditions on the regressor $Z_i$.

# References for Supplemental Material


ANDREWS, D. W. K. AND P. GUGGENBERGER (2009): "Validity of Subsampling and 'Plug-in Asymptotic' Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 25, 669-709.

———(2010): "Asymptotic Size and a Problem with Subsampling and with the m Out of n Bootstrap," *Econometric Theory*, 26, 669-709.

ANDREWS, D. W. K. AND G. SOARES (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica,* 78, 2010, 119-157.

BILLINGSLEY, P. (1995): *Probability and Measure*, 3rd edition. New York: Wiley.

CHERNOZHUKOV, V., S. LEE, AND A. ROZEN (2008): "Inference with Intersection Bounds," unpublished manuscript, Department of Economics, University College London.

LEE, S., K. SONG, AND Y.-J. WHANG (2011): "Testing Functional Inequalities," CEMMAP Working Paper CWP12/11, Institute for Fiscal Studies, University College London.

MANSKI, C. F. AND J. V. PEPPER (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica,* 68, 997-1010.

POLITIS, D. N. AND J. P. ROMANO (1994): "Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions," *Annals of Statistics*, 22, 2031-2050.

POLLARD, D. (1990): *Empirical Process Theory and Application, NSF-CBMS Regional Conference Series in Probability and Statistics*, Vol. II. Institute of Mathematical Statistics.

SCHUMAKER, L. L. (2007): *Spline Functions: Basic Theory,* 3rd edition. Cambridge: Cambridge University Press.